(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 July 2003 (03.07.2003)

PCT

(10) International Publication Number
WO 03/054166 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US02/41225

(22) International Filing Date: 19 December 2002 (19.12.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/342,603 20 December 2001 (20.12.2001) US

(71) Applicant *(for all designated States except US)*: **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **JONES, Karen, Anne** [GB/GB]; 2 Castle Court, Saffron Walden, Essex, Greater London CB10 1BH (GB). **SCHAFER, Alan** [GB/GB]; 38 Canterbury Street, Cambridge, Cambridgeshire CB43QF (GB).

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

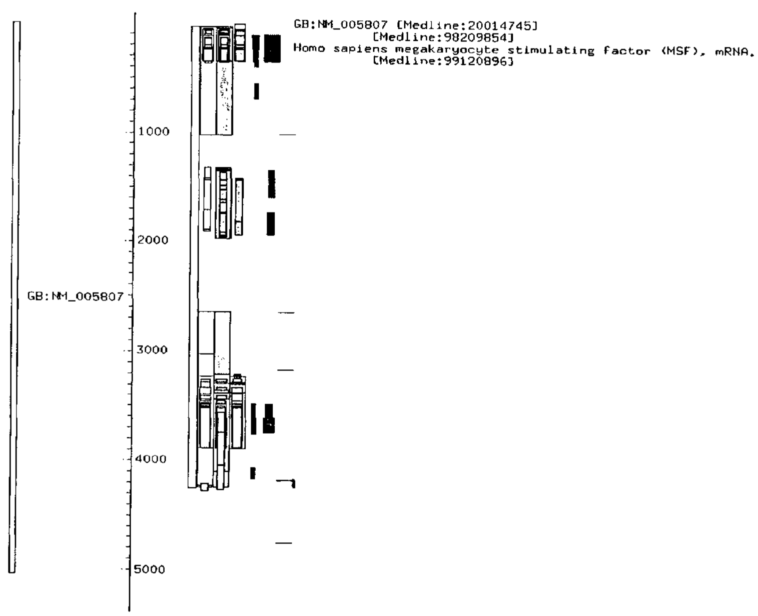
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

[Continued on next page]

(54) Title: NUCLEOTIDE POLYMORPHISMS ASSOCIATED WITH OSTEOARTHRITIS



(57) Abstract: The invention relates to novel polynucleotides associated with human disease, and in particular to osteoarthritis. The invention further relates to polymorphic polynucleotides associated with osteoarthritis. The invention provides methods of determining if a particular polymorphism predisposes an individual to or is associated with the development of osteoarthritis. The invention also provides methods of detecting the presence of one or more polymorphism as an indicator of osteoarthritis, and provides for use of novel polynucleotides of the invention in the development of drugs and in disease treatment.

WO 03/054166 A2

— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

NUCLEOTIDE POLYMORPHISMS ASSOCIATED WITH OSTEOARTHIRTIS

TECHNICAL FIELD

The invention relates in general to polymorphisms in genes associated with osteoarthritis and bone remodeling and methods of identifying individuals having a gene containing a polymorphism associated with osteoarthritis. The invention also relates to a method of detecting an increases susceptibility to a disease in an individual resulting from the presence of a polymorphism or mutation in the gene coding sequence of a osteoarthritis and bone remodeling associated gene.

BACKGROUND OF THE INVENTION

Single nucleotide substitutions and small unique insertions and deletions are the most frequent form of DNA polymorphism and disease-causing mutation in the human genome. These DNA sequence variations, called single nucleotide polymorphisms (SNPs), have gained popularity and have been proposed as the genetic markers of choice for the study of complex genetic traits (Collins et al. 1997 Science 278: 1580- 1581; Risch and Merkangas 1996 Science 273: 1516-1517). Despite the fact that on average approximately one nucleotide position in every 1000 bases along the human chromosome is estimated to differ between any two copies of the chromosome (Cooper et al. 1985 Human Genetics 69: 201-205; Kwok et al. 1996 Genomics 31: 123-126) developing SNP markers is not easy.

It has been suggested that association studies (such as linkage equilibrium studies) with a set of single nucleotide polymorphism (SNP) markers evenly spaced across the genome at approximately 100 KB intervals would provide the necessary power to detect the small effects of each gene involved in a complex trait (Hauser et al. 1996 Genetic Epidemiology 13: 117-137 in Kwok and Chen 1998 Genetic Engineering 20: 125-134, Plenum Press, New York). Alternatively, one can take a candidate gene approach in performing association studies with the use of a set of gene-associated SNP markers to detect these genetic factors (ibid.).

Nucleotide sequence mutations which occur in a gene or gene family, where the gene or gene family is associated with a given disease, may be the basis for susceptibility to or development of the disease.

Arthritis means "inflammation of a joint" and encompasses more than a hundred diseases. They can affect the joints and other connective tissues such as muscles, tendons, ligaments and protective coverings of internal organs. The major arthritis diseases are as follows:

1. osteoarthritis - non-inflammatory degenerative joint disease characterised by splitting and

fragmentation of the articular cartilage, hypertrophy of the bone and changes in the synovial membrane.

2. rheumatoid arthritis - chronic systemic, relapsing disease primarily of the joints which is marked by inflammatory changes in the synovial membranes and adjacent structures.
3. ankylosing spondylitis - inflammatory disease that affects the joints of the lower back which may lead to fusion of the spine
4. gout - caused by formation of uric acid crystals in the joint, leading to inflammation and severe pain.

Osteoarthritis is the most common type of arthritis. It differs from rheumatoid arthritis in that it is primarily a degeneration of the joint tissue that may be accompanied by an inflammatory reaction (Figure 1). Rheumatoid arthritis is an inflammatory disease first and foremost and inflammation of the synovium is the focal point of the disease.

The initiation and progression of osteoarthritis involves multiple pathogenic mechanisms. An imbalance of chondrocyte-controlled anabolic and catabolic processes results in a progressive degradation of the components of the extracellular matrix of the articular cartilage, associated with secondary inflammatory factors. The primary cause of this is unknown but possibly involves a deficiency of cellular response to normal tissue demand or insufficient cellular response to supernormal demand from mechanical loading or injury. The subsequent repair response could induce elevated levels of anabolic molecules, leading to remodelling of the bone and production of osteophytes (bone outgrowths) characteristic of the disease process.

Prevalence and social cost of osteoarthritis.

With approximately 40 million Americans affected by arthritis and other inflammatory diseases, the cost to the healthcare system is significant. Of these 40 million people, 21 million have osteoarthritis and 2.1 million have rheumatoid arthritis. Osteoarthritis is the most common chronic condition and cause of inactivity in patients older than 65. The disease occurs usually at the beginning of the fifth decade of life, with increasing prevalence and incidence with advancing age (Table 2). The prevalence of arthritis is expected to increase by 57% by the year 2020. In the same time period, arthritis-causing activity limitation will increase 66% to 11.6 million people (Lawrence et al 1998). The primary impact of arthritis in the elderly is decreased physical functioning. This can be due to other health-related problems, such as weight gain, cardiovascular disease, GI distress related to treatment, increased psychological distress, decreased social functioning, increased work disability, and

increased healthcare utilization. The current OA treatment, NSAIDs are responsible for the highest number of hospitalisations of any drug category and cause a significant number of internal gastrointestinal bleeding in the elderly population.

The cost of arthritis in the US (including rheumatoid arthritis, osteoarthritis and all other rheumatic conditions) was shown to be $64.8 billion in 1992. Of this, direct costs were an estimated $15.2 billion and indirect costs $49.6 billion (Yelin and Callahan 1995). A 1997 study showed the cost of care for osteoarthritis as $543 per patient per year (Lanes et al 1997). The largest component was hospital care, mostly due to admissions for hip or knee replacement. The cost to the healthcare provider is very high due to the prevalence of the illness.

Unmet medical needs for OA

Current treatment options for osteoarthritis focus on symptom relief whereas truly disease-modifying agents or methods are lacking. Thus, the basic therapy includes common analgesics, nonsteroidal anti-inflammatory drugs, physical therapy, walking aids, and eventually in severe cases, joint replacement surgery. Perhaps because of the difficulties involved in measuring disease progression existing medications do not address the need to prevent further cartilage degradation.

To develop such drugs the following should be in place:

- Compounds that target appropriate biochemical pathways (e.g. Merk's MMP-3 antagonist)
- Clinical studies must be able to measure disease progression in a cost-effective and safe fashion. This could be either an imaging technique or a biomarker that closely correlates with disease progression.
- Disease progression should be detectable within a reasonable time scale (for example, anti-inflammatory clinical studies use the WOMAC pain scale for a period of 6 weeks to measure improvement due to medication).
- The efficacy of the new drug under development should be observable (using either the imaging or biomarker method of assessment) in a sample size comparable to that of other clinical trials.

How can genetics help? Genetic studies have the potential to detect:

- Novel drug targets in the appropriate pathways.
- Individuals with fast progressing osteoarthritis. This would allow a pharmaceutical company to prove efficacy in a relatively small sample size and in a reasonable period

of time, thus cutting costs.

- Reduce variation from biomarker or imaging patterns. For example, let's assume the following response to medication. Although there is a clear patterns of response to medication, it is not statistically significant because of the large amount of variation in disease progression. Lets now assume that there exists a genetic marker that is able to stratify the measurement of disease progression in this hypothetical study. The variance of the marker of disease progression associated with each genotype is smaller than the overall variance. This can be seen as analogous to stratifying a relevant clinical measure in a study (e.g. lipid levels) by gender or by age group. By pooling together both genders or both age groups the variance is larger. If we were now to stratify the results of the previous hypothetical study by genotype we might observe that the therapeutic efficacy is now statistically significant. By stratifying according to genotype it could then be possible to detect statistically significant efficacy in both groups, while meeting the cost and time needs of the entity developing the drug.

Genetic study of osteoarthritis.

Evidence for genetic predisposition to OA.

The nature of the genetic influence in osteoarthritis may involve either a structural defect (that is, collagen), alterations in cartilage or bone metabolism, or a genetic influence on a known risk factor for osteoarthritis such as obesity. Twin studies have show that between 39% and 65% of osteoarthritis in the general population can be attributed to genetic factors (MacGregor and Spector, 1999). Linkage analyses (i.e., common inheritance of affected individuals in the same family) have identified a higher risk ratio for relatives of affected individuals compared to the general population. The power to detect disease-susceptibility loci through linkage analysis using pairs of affected relatives depends on l_R, the risk ratio for type R relatives compared with population prevalence (Risch 1990). Kellgren et al. (1963) compared expected and observed incidence of osteoarthritis in first-degree relatives of probands with multiple osteoarthritis. Based on their results we have estimated l_R for nodal and non-nodal osteoarthritis.

Type of OA	l_R
Nodal (presence of Heberdeen's nodes)	4.5
Non-nodal	4.75

For comparison, concordance for type 2 diabetes ranges between 2-3, and between 4.5 and 5.5 for rheumatoid arthritis. These figures indicate a high genetic component to OA If, however, non-nodal and nodal types of OA are mixed together l_R drops to ~ 2.0 highlighting the importance of careful clinical characterization for genetic studies.

Although it is known that there is a genetic component involved in the etiology of osteoarthritis there is also a need in the art for an improved understanding of the genetic causes of osteoarthritis.

There is also a need in the art for identification of the genes associated with osteoarthritis, and identification of sequence variations in these genes that are associated with osteoarthritis and bone remodeling. The identification of disease related sequence variations in osteoarthritis and bone remodeling associated genes will allow for the development of improved methods of screening for osteoarthritis. These improved screening protocols may be used to identify individuals at high risk for osteoarthritis and in need of preventative treatments.

The identification of disease related sequence variations in osteoarthritis associated genes may facilitate the design of treatment protocols and the identification and design of compounds useful for treatment of osteoarthritis and bone remodeling.

OBJECTS AND SUMMARY OF THE INVENTION

An object of the present invention is to provide candidate genes associated with osteoarthritis and bone remodeling.

It is another object of the present invention to provide a variant nucleotide in a candidate gene associated with osteoarthritis and bone remodeling.

Another object of the present invention is to provide methods of detecting variant nucleotides in a gene in individuals at risk for osteoarthritis.

Another object of the present invention is to provide methods of determining if a variant nucleotide is associated with a predisposition to osteoarthritis.

Another object of the present invention is to provide candidate genes associated with the osteoarthritis and bone remodeling.

The invention further comprises isolated polynucleotides which contain the single nucleotide polymorphisms selected from the Sequence Listing, or its perfect complement.

The invention further comprises an isolated polynucleotide segment of between 10 and 100 bases of which 10 contiguous bases including a polymorphic site are from a sequence selected from the Sequence Listing, or its perfect complement.

The invention further comprises a probe or target sequence used for genotyping where the

probe or target sequence has at least 10 contiguous bases containing a polymorphic site identified and from a sequence selected from the Sequence Listing, or its perfect complement.

The invention further comprises a method for determining a base occupying a polymorphic site in a nucleic acid comprising obtaining the nucleic acid in a sample from an individual or plurality of individuals and determining a base occupying a polymorphic site in a sequence selected from the group consisting of the Sequence Listing and their perfect complements which occurs in the sample nucleic acid.

DESCRIPTION OF THE COMPACT DISK-RECORDABLES (CD-R)

CD-R (Copy 1)contains the Sequence Listing formatted in plain ASCII text and Tables 1 and 2. CD-R (Copy 1) is labeled with Identification No. GX-0022P-1.

CD-R (Copy 2) is an exact copy of CD-R (Copy 1). CD-R (Copy 2) is labeled with Identification No. GX-0022-1 P (Copy 2).

CD-R (Copy 3) contains the Computer Readable Form of the Sequence Listing in compliance with 37 C.F.R. §1.821(e), and specified by 37 C.F.R. §1.824. CD-R (Copy 3) is labeled with Identification No. GX-0022-1 P (Copy 3).

The material on CD-R 1, 2 and 3 is incorporated by reference into the specification.

BRIEF DESCRIPTION OF THE TABLES AND DRAWINGS

These and other features, aspects, and advantages of the present invention will become better understood with regard to the following description, appended claims, and accompanying tables drawings where:

Table 1 presents the genomic or cDNA structure of osteoarthritis candidate gene sequences and the identity and position of polymorphisms which are the subject of the invention. This table has the form wherein:

a. The DNA change given for an allele is not strand specific; it can be on either strand of the DNA molecule.

b. Single Nucleotide Polymorphisms can be recorded as IUPAC ambiguity symbols, as follows:

M	A or C
R	A or G
S	C or G
K	G or T

W A or T

Y C or T

c. Other allele types, such as insertions and deletions, are given in the form: ACA>AA or AA>ACA and in such cases the coordinates of the allele include the two invariant flanking bases.

d. DNA sequence names are of the form: XX:IIIIII[_VV], where XX gives the database of origin, as follows:

EM EMBL

FN Incyte FL sequence read

GB GenBank

IN Incyte proprietary sequence

LG LifeSeq Gold gene template

IIIIII gives the sequence ID or accession number for the sequence. In most cases if it is an accession number it will be followed by _VV where VV is the sequence version in the EMBL or GenBank database.

e. The overall structure of a record in the patent structure is described as follows. Items in {braces} indicate a field that is filled in. Items in [square brackets] may or may not be present. These entries define a larger virtual sequence - a "link" - composed of real database subsequences. Alleles are annotated onto real sequences, and genomic structure onto the link.

{Locus ID}

[Full name : {full name}]

Link : {link name}

Subsequence {name} {link start position} {link stop position} {SEQ ID NO}

[...]

CDS {name} {SEQ ID NO}

exon/ORF {link start position} {link stop position}

[...]

[...]

Allele {seq name} {SEQ ID NO} {seq start} {seq stop} {dna change}

source {original SNP data source} {SNP id in that source}

[...]

consequence {CDS name} {CDS SEQ ID NO} {class} [{peptide pos} {peptide

change}]

[...]

[...]

f. Sources. SNPs may have been noted in one of several sources:

dbSNP The NCBI public dbSNP databank

isSNP In silico SNPs from LifeSeq sequence assembly.

wetSNP Alleles determined by SSCP.

Alleles which have a wetSNP entry are experimentally verified. Alleles which are isSNP and/or dbSNP only are predictions by computer software of where these SNPs map to, and are *not* experimentally verified.

g. Consequences

The classes of consequence are as follows:

Silent The allele does not cause a peptide change

Missense The allele causes an amino acid substitution

Frameshift The allele causes a frame shift in the CDS

Intron The allele lies wholly within an intron

5' The allele lies 5' of the CDS

3' The allele lies 3' of the CDS

Unknown The consequence is undefined - for example the allele straddles an intron/exon boundary.

Silent and Missense consequences also supply details of the amino acid position of the change, and prediction of what the affected amino acid is, and what it is substituted to. There may be multiple consequence lines if the locus contains multiple CDS forms.

h. Sequence and exon positions

Sequence coordinates are always given on the forward strand of the link. Therefore, if a sequence or exon is actually on the reverse strand of the link, its start position will be larger than its stop position.

i. Exon order in CDS definitions

The exons are given in 5' to 3' order. Consequently, reverse strand CDS start from high coordinate numbers downwards.

j. Link object types

Loci may have more than one link object, composed of different DNA sequences. Typically there might be one genomic and one cDNA link object.

Table 2 presents the population frequency of polymorphisms in the candidate genes and summarizes various information from Table 2 relating to the polymorphism.

Figure 1 illustrates the cDNA structure of the locus and relative positions of identified SNPs for megakaryocyte stimulating factor (MSF).

Figure 2 illustrates the genomic structure of the locus, exons composing multiple CDS, and relative positions of identified SNPs for megakaryocyte stimulating factor (MSF).

The figures show (from left to right) the real sequences making up the linked genomic structure for the locus, a scale in link coordinates (negative numbers would indicate a view of the reverse strand), one or more CDSs representing the positions of exons, horizontal bars representing the positions of identified SNPs (alleles) from the various sources, and shaded boxes showing regions targeted for screening by SSCP.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

Before the present compositions and methods are described, it is understood that embodiments of the invention are not limited to the particular machines, instruments, materials, and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the invention.

As used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a nucleic acid probe" includes a plurality of such nucleic acid probes, and a reference to "a gene" is a reference to one or more genes and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with various embodiments of the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

Definitions

As used herein, "polymorphism" refers to a nucleotide alteration that either predisposes an individual to a disease or is not associated with a disease, which occurs as a result of a substitution, insertion or deletion.

More particularly, a "polymorphism" or "polymorphic variation" may be a nucleic acid sequence variation, as compared to the naturally occurring sequence, resulting from either a nucleotide deletion, an insertion or addition, or a substitution, which is present at a frequency of greater than 1% in a population.

As used herein, "neutral polymorphism" refers to a polymorphism which is present at a frequency of greater than 1% in a population, which does not alter gene function or phenotype, and thus is not associated with a predisposition to or development of a disease.

As used herein "polynucleotide sequence" refers to a sense or antisense nucleic acid sequence comprising RNA, cDNA, genomic DNA, synthetic forms and mixed polymers, that may be chemically or biochemically modified or may contain non-natural or derivatized nucleotide bases.

As used herein "mutation" refers to a variation in the nucleotide sequence of a gene or regulatory sequence as compared to the naturally occurring or normal nucleotide sequence. A mutation may result from the deletion, insertion or substitution of more than one nucleotide (e.g., 2, 3, 4, or more nucleotides) or a single nucleotide change such as a deletion, insertion or substitution. The term "mutation" also encompasses chromosomal rearrangements.

As used herein, "nucleic acid probe" refers to an oligonucleotide, nucleotide or polynucleotide, and fragments and portions thereof, and to DNA or RNA of genomic or synthetic origin which may be single- or double- stranded, which represents the sense or antisense strand. Both terms "nucleic acid probe" and "DNA fragment" refer to a length of polynucleotide, for example, as small as 5 nucleotides, 10, 20, 25, 40, 50, 75, 100, 250, 400, 500 and 1 kb, and as large as 5-10kb.

As used herein, "alteration" refers to a change in either a nucleotide or amino acid sequence, as compared to the naturally occurring sequence, resulting from a deletion, an insertion or addition, or a substitution.

As used herein, "deletion" refers to a change in either nucleotide or amino acid sequence wherein one or more nucleotides or amino acid residues, respectively, are absent.

As used herein, "insertion" or "addition" refers to a change in either nucleotide or amino acid sequence wherein one or more nucleotides or amino acid residues, respectively, have been added.

As used herein, "substitution" refers to a replacement of one or more nucleotides or amino acids by different nucleotides or amino acid residues, respectively.

As used herein, "specifically hybridizable" refers to a nucleic acid or fragment thereof that

hybridizes to another nucleic acid (or a complementary strand thereof) due to the presence of a region that is at least approximately 90% homologous, preferably at least approximately 90-95% homologous, and more preferably approximately 98-100% homologous, as are polynucleotides that hybridize to a partner under stringent hybridization conditions. "Stringent" hybridization conditions are defined hereinbelow for various hybridization protocols. A probe that is specifically hybridizable to a given sequence can be used to detect a 1 bp out of 10 bp (10%) or a 1 bp out of 2O bp (5%) difference between nucleic acid sequences and is therefore useful for discriminating between a wild type and a mutant form of a gene of interest.

As used herein, "amino acid sequence" refers to the sequential array of amino acids that have been joined by peptide bonds between the carboxylic acid group of one amino acid and the amino group of the adjacent amino acid to form long linear polymers comprising proteins.

As used herein, "amino acid" refers to protein subunit molecules that contain a carboxylic acid group, and an amino group, both linked to a single carbon atom.

A polypeptide is said to be "encoded" by a polynucleotide if the polynucleotide, either in its native state or in a recombinant form can be transcribed and/or translated to produce the mRNA for and/or the polypeptide or a fragment thereof.

As used herein, "gene " refers to a region of DNA which includes a portion which can be transcribed into RNA, and which may contain an open reading frame or coding region (also referred to as an exon) which encodes a protein, a non-coding region (also referred to as an intron), and a specific regulatory region comprising the DNA regulatory elements which control expression of the transcribed region.

As used herein, "coding region" refers to a region of DNA which encodes a protein, also known as an exon.

As used herein, "non-coding region" refers to a region of DNA which does not encode a protein coding region, also known as an intron, and is not included in the RNA molecule that is synthesized from a particular gene.

As used herein, "regulatory region" refers to DNA sequences which are located either 5' of the transcription start site, 3' or the transcription termination site, within an intron or exon, capable of ensuring that the gene is transcribed at the proper time and in the appropriate cell type.

As used herein, "consensus DNA sequence" or "wild-type DNA sequence" refers to a sequence wherein every position represents the nucleotide that occurs with the highest frequency when many actual sequences are compared. As used herein, "consensus DNA sequence" or "wild--type DNA sequence" also refers to the normal, naturally occurring DNA sequence.

As used herein, a given sequence (or mutation or polymorphism) "associated with" osteoarthritis refers to a nucleic acid sequence that increases susceptibility to the disease, predisposes an individual to the disease or contributes to the disease, wherein the nucleic acid sequence is present at a higher frequency (at least 5%, preferably 10%, more preferably 25% higher) in individuals with the disease as compared to individuals who do not have the disease.

As used herein, a sequence "not associated with" osteoarthritis refers to a nucleic acid sequence that does not increase susceptibility to the disease, predispose an individual to the disease or contribute to the disease, wherein the nucleic acid sequence is not present at a higher frequency in individuals with the disease, and thus is present at a frequency about equal to its frequency in individuals who do not have the disease.

As used herein, "amplifying" refers to producing additional copies of a nucleic acid sequence, preferably by the method of polymerase chain reaction (Mullis and Faloona, 1987, Methods Enzymol, 155: 335).

As used herein, "oligonucleotide primers" refer to single stranded DNA or RNA molecules that are hybridizable to a nucleic acid template and prime enzymatic synthesis of a second nucleic acid strand. Oligonucleotide primers useful according to the invention are between 5 to 100 nucleotides in length, preferably 20-60 nucleotides in length, and more preferably 20-40 nucleotides in length.

As used herein, "sequencing" refers to determining the precise nucleotide composition or sequence of a nucleic acid region by methods well known in the art (see Ausubel et al., supra and Sambrook et al., supra).

As used herein, "comparing" a sequence refers to determining if the nucleotides at one or more positions in a particular region of a nucleic acid fragment are identical for any two or more sequences. According to the invention, sequence comparisons can be performed by using computer program analysis as described below in Section F entitled "Identification and Characterization of Polymorphisms".

As used herein, "sequence differences" or "sequence variations" refer to nucleotide changes, at one or more positions between any two or more sequences being compared.

As used herein, "determining the presence of polymorphic variations" refers to using methods well known in the art to identify a nucleotide, at one or more positions within a particular nucleic acid region, that is distinct from the nucleotide present in the naturally occurring, wild-type or consensus sequence, resulting from either a nucleotide deletion, an insertion or addition, or a substitution.

As used herein, "determining the absence of polymorphic variations" refers to using methods well known in the art to determine that the nucleotides present at every position analyzed in a

particular nucleic acid region are identical to the nucleotides present in the naturally occurring, wild-type or consensus sequence.

As used herein, "genotyping" refers to determining the composition of the genetic material that is inherited by an organism from its parents.

As used herein, "biological sample" refers to a tissue or fluid sample containing a polynucleotide or polypeptide of interest, and isolated from an individual including but not limited to plasma, serum, spinal fluid, lymph fluid, urine, stool, external secretions of the skin, respiratory, intestinal and genitoruinary tracts, saliva, blood cells, tumors, organs, tissue and samples of *in vitro* cell culture constituents.

As used herein, "amplimers" refer to a specific fragment of DNA generated by PCR that is at least 30 bp in length and is preferably between 50 and lOObp in length, and is more preferably between 150-300bp in length, with a melting temperature in the range of approximately 60-62^{0}C.

As used herein, "phenotype" refers to the biological appearances of an organism or a tissue derived from an organism, wherein biological appearances include chemical, structural and behavioral attributes, and excludes genetic constitution.

As used herein, "genotype" refers to the genetic material that is inherited by an organism from its parents.

As used herein, "genetic susceptibility to osteoarthritis" refers to an increased risk of developing osteoarthritis resulting from specific DNA differences relative to non-susceptible individuals. Preferably an individual who is genetically susceptible to osteoarthritis has a 5-100%, and more preferably a 25-50% greater chance of developing osteoarthritis, as compared to non-susceptible individuals.

As used herein, "diagnostic" refers to the practice of identifying a disease from the signs and symptoms of an individual including the DNA sequences of genes that are associated with an increased susceptibility to the disease. "Diagnostic" also refers to the practice of stratifying patient populations based on the efficacy or toxicity of a composition, and the predictive placement of an individual in a response strata based on stata-associated parameters.

As used herein, "prognosis" refers to the possibility of recovering from a particular disease or condition, and also refers to risk assessment of developing a particular disease or condition.

THE INVENTION

Various embodiments of the invention include polynucleotides and polymorphic polynucleotides associated with a given human disease, for example, with osteoarthritis. The invention also provides a

gene sequence containing one or more polymorphic nucleotides associated with a predisposition to or the development of a given human disease such as osteoarthritis. The invention also relates to polypeptides encoded by the polynucleotides or the polymorphism-containing gene. The invention also provides methods of detecting a polymorphism according to the invention in individuals at risk for osteoarthritis, and for determining if a given polymorphism is associated with a predisposition to the disease. The invention also discloses polymorphism(s) that are either associated with or are not associated with (i.e., are neutral) osteoarthritis. A polymorphism in a given gene can be utilized in various diagnostic and therapeutic methods and procedures, for example, in nucleic acid and peptide diagnosis, drug screening and design, and in gene and peptide therapy. A polymorphism associated with a given gene can be utilized in various gene expression systems and assays designed to analyze gene regulation and expression.

A. Design and Synthesis of Oligonucleotide Primers

According to the present invention, oligonucleotide primers are disclosed that are useful for determining the sequence of a particular allele of a gene. The invention also discloses oligonucleotide primers designed to amplify a region of a gene that is known to contain a polymorphism. The invention also discloses oligonucleotide primers designed to anneal specifically to a particular allele of a gene.

Oligonucleotide primers useful according to the invention are single-stranded DNA or RNA molecules that are hybridizable to a nucleic acid template and prime enzymatic synthesis of a second nucleic acid strand. The primer is complementary to a portion of a target molecule present in a pool of nucleic acid molecules. It is contemplated that oligonucleotide primers according to the invention are prepared by synthetic methods, either chemical or enzymatic. Alternatively, such a molecule or a fragment thereof is naturally-occurring, and is isolated from its natural source or purchased from a commercial supplier. Oligonucleotide primers are 5 to 100 nucleotides in length, ideally from 20 to 40 nucleotides, although oligonucleotides of different length are of use.

Pairs of single-stranded DNA primers can be annealed to sequences within or surrounding a gene on chromosome Y in order to prime amplifying DNA synthesis of a region of a gene. A complete set of gene primers will allow synthesis of all of the nucleotides of the coding sequences, e.g., the exons, introns and control regions. Preferably, the set of primers will also allow synthesis of both intron and exon sequences.

Allele-specific primers are also useful, according to the invention. Such primers will anneal only to a particular-mutant allele (e.g. alleles containing a polymorphism), and thus will only amplify a product if the template also contains the polymorphism. Allele specific primers that anneal only to a

wild type gene sequence are also useful according to the invention.

Typically, selective hybridization occurs when two nucleic acid sequences are substantially complementary (at least about 65% complementary over a stretch of at least 14 to 25 nucleotides, preferably at least about 75%, more preferably at least about 90% complementary). See Kanehisa, M., 1984, Nucleic Acids Res. 12: 203, incorporated herein by reference. As a result, it is expected that a certain degree of mismatch at the priming site is tolerated. Such mismatch may be small, such as a mono-, di- or tri-nucleotide. Alternatively, it may encompass loops, which are defined as regions in which there exists a mismatch in an uninterrupted series of four or more nucleotides.

Numerous factors influence the efficiency and selectivity of hybridization of the primer to a second nucleic acid molecule. These factors, which include primer length, nucleotide sequence and/or composition, hybridization temperature, buffer composition and potential for steric hindrance in the region to which the primer is required to hybridize, will be considered when designing oligonucleotide primers according to the invention.

A positive correlation exists between primer length and both the efficiency and accuracy with which a primer will anneal to a target sequence. In particular, longer sequences have a higher melting temperature (T_M) than do shorter ones, and are less likely to be repeated within a given target sequence, thereby minimizing promiscuous hybridization. Primer sequences with a high G-C content or that comprise palindromic sequences tend to self-hybridize, as do their intended target sites, since unimolecular, rather than bimolecular, hybridization kinetics are generally favored in solution. However, it is also important to design a primer that contains sufficient numbers of G-C nucleotide pairings since each G-C pair is bound by three hydrogen bonds, rather than the two that are found when A and T bases pair to bind the target sequence, and therefore forms a tighter, stronger bond. Hybridization temperature varies inversely with primer annealing efficiency, as does the concentration of organic solvents, e.g. formamide, that might be included in a priming reaction or hybridization mixture, while increases in salt concentration facilitate binding. Under stringent annealing conditions, longer hybridization probes (of use, for example, in Northern analysis), or synthesis primers, hybridize more efficiently than do shorter ones, which are sufficient under more permissive conditions. Stringent hybridization conditions typically include salt concentrations of less than about lM, more usually less than about 500 mM and preferably less than about 200 mM. Hybridization temperatures range from as low as 0^0C to greater than 22^0C, greater than about 30^0C, and (most often) in excess of about 37^0C. Longer fragments may require higher hybridization temperatures for specific hybridization. As several factors affect the stringency of hybridization, the combination of parameters is more important than the absolute measure of a single factor.

Oligonucleotide primers can be designed with these considerations in mind and synthesized according to the following methods.

1. Oligonucleotide Primer Design Strategy

The design of a particular oligonucleotide primer for the purpose of sequencing or PCR involves selecting a sequence that is capable of recognizing the target sequence, but has a minimal predicted secondary structure. The oligonucleotide sequence binds only to a single site in the target nucleic acid. Furthermore, the Tm of the oligonucleotide is optimized by analysis of the length and GC content of the oligonucleotide. Furthermore, when designing a PCR primer useful for the amplification of genomic DNA, the selected primer sequence does not demonstrate significant matches to sequences in the GenBank database (or other available databases).

The design of a primer is facilitated by the use of readily available computer programs, developed to assist in the evaluation of the several parameters described above and the optimization of primer sequences. Examples of such programs are "PrimerSelect" of the DNAStar™ software package (DNAStar, Inc.; Madison, WI), OLIGO 4.0 (National Biosciences, Inc.), PRIMER, Oligonucleotide Selection Program, PGEN and Amplify (described in Ausubel et al., 1995, Short Protocols in Molecular Biology, 3rd Edition, John Wiley & Sons). Primers are designed with sequences that serve as targets for other primers to produce a PCR product that has known sequences on the ends which serve as targets for further amplification (e.g. to sequence the PCR product). If many different genes are amplified with specific primers that share a common 'tail' sequence', the PCR products from these distinct genes can subsequently be sequenced with a single set of primers. Alternatively, in order to facilitate subsequent cloning of amplified sequences, primers are designed with restriction enzyme site sequences appended to their 5' ends. Thus, all nucleotides of the primers are derived from gene sequences or sequences adjacent to a gene, except for the few nucleotides necessary to form a restriction enzyme site. Such enzymes and sites are well known in the art. If the genomic sequence of a gene and the sequence of the open reading frame of a gene are known, design of particular primers is well within the skill of the art.

2. Synthesis

The primers themselves are synthesized using techniques which are also well known in the art. Once designed, oligonucleotides are prepared by a suitable method, e.g. the phosphoramidite method described by Beaucage and Carruthers (1981, Tetrahedron Lett., 22:1859) or the triester method according to Matteucci et al. (1981, J. Am. Chem. Soc., 103:3185), both incorporated herein by reference, or by other chemical methods using either a commercial automated oligonucleotide

synthesizer (which is commercially available) or VLSIPS™ technology.

B. Production of a Polynucleotide Sequence

The invention discloses polynucleotide sequences comprising polymorphisms. The polynucleotide sequences of the invention are specifically hybridizable to a mutant form of a gene and are therefore useful for discriminating between a wild-type form of a gene and a mutant form of a gene. The polynucleotide sequences of the invention may also be useful for expression of the encoded protein or a fragment thereof. The invention also features antisense polynucleotide sequences complementary to polynucleotide sequences comprising polymorphisms. Antisense polynucleotide sequences are useful according to the invention for inhibiting expression of an allelic form of a gene.

The present invention utilizes polynucleotide sequences and fragments comprising RNA, cDNA, genomic DNA, synthetic forms, and mixed polymers. The invention includes both sense and antisense strands of the polynucleotide sequences. According to the invention, the polynucleotide sequences may be chemically or biochemically modified or may contain non-natural or derivatized nucleotide bases. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (e.g. methyl phosphonates, phosphorodithioates. etc.), pendent moieties (e.g., polypeptides), intercalators, (e.g. acridine, psoralen, etc.) chelators, alkylators, and modified linkages (e.g. alpha anomeric nucleic acids, etc.) Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

The polynucleotide may be a naturally occurring polynucleotide, or may be a structurally related variant of such a polynucleotide having modified bases and/or sugars and/or linkages. The term "polynucleotide" as used herein is intended to cover all such variants.

Modifications, which may be made to the polynucleotide may include (but are not limited to) the following types:

a) <u>Backbone modifications</u>

i) phosphorothioates (X or Y or W or Z = S or any combination of two or more with the remainder as 0).

e.g. Y=S (Stein et al., 1988, <u>Nucleic Acids Res.</u>, 15:3209), X=S (Cosstick and Vyle, 1989, <u>Tetrahedron Letters</u>, 30:4693), Y and Z=S (Brill et al., 1989, <u>J. Amer. Chem. Soc.</u>, 111:2321)

ii) methylphosphonates (eg Z=methyl (Miller et al., 1980, <u>J. Biol. Chem.</u>, 255:9569))

iii) phosphoramidates (Z = N-(alkyl)$_2$ e.g. alkyl methyl, ethyl. butyl) (Z=morpholine or piperazine) (Agrawal et al., 1988, Proc. Natl. Acad. Sci.. USA, 85;7079) (X or W = NH) (Mag and Engels. 1988, Nucleic Acids Res., 16:3525)

iv) phosphotriesters (Z=O-alkyl e.g. methyl, ethyl etc) (Miller et al., Biochemistry, 21:5468)

v) phosphorus-free linkages (e.g. carbamate, acetamidate, acetate) (Gait et al., 1974, J Chem.Soc. Perkin I, 1684, Gait et al., 1979, J Chem.Soc. Perkin I, 1389)

b) Sugar modifications

i) 2'-deoxynucleosides (R=H)

ii) 2'-O-methylated nucleosides (R=OMe) (Sproat et al., 1989, Nucleic Acids Res., 17: 3373)

iii) 2'-fluoro-2'-deoxynucleosides (R=F) (Krug et al., 1989, Nucleosides and Nucleotides, 8:1473)

c) Base modifications - (for a review see Jones, 1979, Int. J. Biolog. Macromolecules, 1:194)

i) pyrimidine derivatives substituted in the 5-position (e.g. methyl, bromo, fluoro etc) or replacing a carbonyl group by an amino group (Piccirilli et al., 1990, Nature, 343:33).

ii) purine derivatives lacking specific nitrogen atoms (e.g. 7-deaza adenine, hypoxanthine) or functionalized in the 8-position (e.g. 8-azido adenine, 8-bromo adenine)

d) Polynucleotides covalently linked to reactive functional groups, e.g.:

i) psoralens (Miller et al., 1988, Nucleic Acids Res. Special Pub. No. 20:113, phenanthrolines (Sun et al., 1988, Biochemistry, 27:6039), mustards (Vlassov et al., 1988, Gene, 72:313) (irreversible cross-linking agents with or without the need for co-reagents)

ii) acridine (intercalating agents) (Helene et al., 1985, Biochimie, 67:777)

iii) thiol derivatives (reversible disulphide formation with proteins) (Connolly and Newman, 1989, Nucleic Acids Res., 17:4957)

iv) aldehydes (Schiffs base formation)

v) azido, bromo groups (UV cross-linking)

vi) ellipticines (photolytic cross-linking) (Perrouault et al., 1990, Nature, 344:358)

e) Polynucleotides covalently linked to lipophilic groups or other reagents capable of improving uptake by cells, e.g.:

i) cholesterol (Letsinger et al., 1989, Proc. Natl. Acad. Sci. USA, 86:6553), polyamines (Lemaitre et al., 1987, Proc. Natl. Acad. Sci. USA, 84: 648), other soluble polymers (e.g. polyethylene glycol)

f) Polynucleotides containing alpha-nucleosides (Morvan et al., Nucleic Acids Res., 15: 3421)

g) Combinations of modifications a)-f)

It should be noted that such modified polynucleotides, while sharing features with polynucleotides designed as "anti-sense" inhibitors, are distinct in that the compounds correspond to sense-strand sequences and the mechanism of action depends on protein-nucleic acid interactions and does not depend upon interactions with nucleic acid sequences.

1. Polynucleotide Sequences Comprising DNA

a. Cloning

Polynucleotide sequences comprising DNA can be isolated from cDNA or genomic libraries (including YAC and BAC libraries) by cloning methods well known to those skilled in the art (Ausubel et al., supra). Briefly, isolation of a DNA clone comprising a particular polynucleotide sequence involves screening a recombinant DNA or cDNA library and identifying the clone containing the desired sequence. Cloning will involve the following steps. The clones of a particular library are spread onto plates, transferred to an appropriate substrate for screening, denatured, and probed for the presence of a particular sequence. A description of hybridization conditions, and methods for producing labeled probes is included below.

The desired clone is preferably identified by hybridization to a nucleic acid probe or by expression of a protein that can be detected by an antibody. Alternatively, the desired clone is identified by polymerase chain amplification of a sequence defined by a particular set of primers according to the methods described below.

The selection of an appropriate library involves identifying tissues or cell lines that are an abundant source of the desired sequence. Furthermore, if the polynucleotide sequence of interest contains regulatory sequence or intronic sequence a genomic library is screened (Ausubel et al., supra).

b. Genomic DNA

Polynucleotide sequences of the invention are amplified from genomic DNA. Genomic DNA is isolated from tissues or cells according to the following method.

To facilitate detection of a variant form of a gene from a particular tissue, the tissue is isolated free from surrounding normal tissues. To isolate genomic DNA from mammalian tissue, the tissue is minced and frozen in liquid nitrogen. Frozen tissue is ground into a fine powder with a prechilled mortar and pestle, and suspended in digestion buffer (100 mM NaCl, 10 mM TrisCl, pH 8.0, 25 mM EDTA, pH 8.0, 0.5% (w/v) SDS, 0.1 mg/ml proteinase K) at 1.2ml digestion buffer per 100mg of tissue. To isolate genomic DNA from mammalian tissue culture cells, cells are pelleted by

centrifugation for 5 min at 500 x g, resuspended in 1-10 ml ice-cold PBS, repelleted for 5 min at 500 x g and resuspended in 1 volume of digestion buffer.

Samples in digestion buffer are incubated (with shaking) for 12-18 hours at 50^{0}C, and then extracted with an equal volume of phenol/chloroform/isoamyl alcohol. If the phases are not resolved following a centrifugation step (10 min at 1700 x g), another volume of digestion buffer (without proteinase K) is added and the centrifugation step is repeated. If a thick white material is evident at the interface of the two phases, the organic extraction step is repeated. Following extraction the upper, aqueous layer is transferred to a new tube to which will be added 1/2 volume of 7.5M ammomum acetate and 2 volumes of 100% ethanol. The nucleic acid is pelleted by centrifugation for 2 min at 1700 x g, washed with 70% ethanol, air dried and resuspended in TE buffer (10 mM TrisCl, pH 8.0, 1 mM EDTA, pH 8.0) at lmg/ml. Residual RNA is removed by incubating the sample for 1 hour at 37^{0}C in the presence of 0.1% SDS and 1 mg/ml DNAse-free RNASE, and repeating the extraction and ethanol precipitation steps. The yield of genomic DNA, according to this method is expected to be approximately 2 mg DNA/1 g cells or tissue (Ausubel et al., supra). Genomic DNA isolated according to this method can be used for Southern blot analysis, restriction enzyme digestion, dot blot analysis or PCR analysis, according to the invention.

c. Restriction digest (of cDNA or genomic DNA)

Following the identification of a desired cDNA or genomic clone containing a particular sequence, polynucleotides of the invention are isolated from these clones by digestion with restriction enzymes.

The technique of restriction enzyme digestion is well known to those skilled in the art (Ausubel et al., supra). Reagents useful for restriction enzyme digestion are readily available from commercial vendors including New England Biolabs, Boebringer Mannheim, Promega, as well as other sources.

d. PCR

Polynucleotide sequences of the invention are amplified from genomic DNA or other natural sources by the polymerase chain reaction (PCR). PCR methods are well-known to those skilled in the art.

PCR provides a method for rapidly amplifying a particular DNA sequence by using multiple cycles of DNA replication catalyzed by a thermostable, DNA-dependent DNA polymerase to amplify the target sequence of interest. PCR requires the presence of a nucleic acid to be amplified, two single stranded oligonucleotide primers flanking the sequence to be amplified, a DNA polymerase, deoxyribonucleoside triphosphates, a buffer and salts.

The method of PCR is well known in the art. PCR, is performed as described in Mullis and

Faloona, 1987, Methods Enzymol., 155: 335, herein incorporated by reference.

PCR is performed using template DNA (at least 1 fg; more usefully, 1 - 1000 ng) and at least 25 pmol of oligonucleotide primers. A typical reaction mixture includes: 2 ml of DNA, 25 pmol of oligonucleotide primer, 2.5 ml of lOx PCR buffer 1 (Perkin-Elmer, Foster City, CA), 0.4 ml of 1.25 mM dNTP, 0.15 ml (or 2.5 units) of Taq DNA polymerase (Perkin Elmer, Foster City, CA) and deionized water to a total volume of 25 ml. Mineral oil is overlaid and the PCR is performed using a programmable thermal cycler.

The length and temperature of each step of a PCR cycle, as well as the number of cycles, are adjusted according to the stringency requirements in effect. Annealing temperature and timing are determined both by the efficiency with which a primer is expected to anneal to a template and the degree of mismatch that is to be tolerated. The ability to optimize the stringency of primer annealing conditions is well within the knowledge of one of moderate skill in the art. An annealing temperature of between 30°C and 72°C is used. Initial denaturation of the template molecules normally occurs at between 92°C and 99°C for 4 minutes, followed by 20-40 cycles consisting of denaturation (94-99°C for 15 seconds to 1 minute), annealing (temperature determined as discussed above; 1-2 minutes), and extension (72°C for 1 minute). The final extension step is generally carried out for 4 minutes at 72°C, and may be followed by an indefinite (0-24 hour) step at 4°C.

Several techniques for detecting PCR products quantitatively without electrophoresis may be useful according to the invention in order to make it more suitable for easy clinical use. One of these techniques, for which there are commercially available kits such as Taqman™ (Perkin Elmer, Foster City, CA), is performed with a transcript-specific antisense probe. This probe is specific for the PCR product (e.g. a nucleic acid fragment derived from a gene) and is prepared with a quencher and fluorescent reporter probe complexed to the 5' end of the oligonucleotide. Different fluorescent markers can be attached to different reporters, allowing for measurement of two products in one reaction. When Taq DNA polymerase is activated, it cleaves off the fluorescent reporters of the probe bound to the template by virtue of its 5'-to-3' nucleolytic activity. In the absence of the quenchers, the reporters now fluoresce. The color change in the reporters is proportional to the amount of each specific product and is measured by a fluorometer; therefore, the amount of each color can be measured and the PCR product can be quantified. The PCR reactions can be performed in 96 well plates so that samples derived from many individuals can be processed and measured simultaneously. The Taqman™ system has the additional advantage of not requiring gel electrophoresis and allows for quantification when used with a standard curve.

2. Polynucleotide Sequences Comprising RNA

The present invention also provides a polynucleotide sequence comprising RNA. A polynucleotide comprising RNA is useful for detecting snps and polymorphisms by techniques including but not limited to hybridization methods or the RNase protection method. A polynucleotide comprising RNA is also useful as a template for the *in vitro* production of protein. A polynucleotide comprising RNA is also useful for detecting and localizing specific mRNA sequences by *in situ* hybridization.

Polynucleotide sequences comprising RNA can be produced according to the method of *in vitro* transcription.

The technique of *in vitro* transcription is well known to those of skill in the art. Briefly, the gene of interest is inserted into a vector containing an SP6, T3 or T7 promoter. The vector is linearized with an appropriate restriction enzyme that digests the vector at a single site located downstream of the coding sequence. Following a phenol/chloroform extraction, the DNA is ethanol precipitated, washed in 70% ethanol, dried and resuspended in sterile water. The *in vitro* transcription reaction is performed by incubating the linearized DNA with transcription buffer (200 mM TrisCl, pH 8.0,40 mM $MgCl_2$, 10 mM spermidine, 250 NaCl [T7 or T3] or 200 mM TrisCl, pH 7.5,30 mM $MgCl_2$, 10mM spermidine [SP6]), dithiothreitol, RNASE inhibitors, each of the four ribonucleoside triphosphates, and either SP6, T7 or T3 RNA polymerase for 30 min at 37^{0}C. To prepare a radiolabeled polynucleotide comprising RNA, unlabeled UTP will be omitted and ~SUTP will be included in the reaction mixture. The DNA template is then removed by incubation with DNaseI. Following ethanol precipitation, an aliquot of the radiolabeled RNA is counted in a scintillation counter to determine the cpm/ml (Ausubel et al., supra).

Alternatively, polynucleotide sequences comprising RNA are prepared by chemical synthesis techniques such as solid phase phosphoramidite (described above).

3. Polynucleotide Sequences Comprising Oligonucleotides

A polynucleotide sequence comprising oligonucleotides can be made by using oligonucleotide synthesizing machines which are commercially available (described above).

4. Polynucleotide Sequences Encoding Fusion Proteins

Polynucleotide sequences of the invention can be used to express the protein product (or fragment thereof) of the gene of interest by inserting the polynucleotide sequence into an expression vector. Expression vectors suitable for protein expression in mammalian cells, bacterial cells, insect cells or plant cells are well known in the art and are described in Section H entitled "Production of a Mutant Protein".

Polynucleotide sequences of the invention can be used to prepare hybrid polynucleotides

comprising a sequence of a gene adjacent to a sequence encoding a foreign protein or a fragment thereof (e.g lacZ, trpE, glutathionine S-transferase or thioredoxin) or a protein tag (hemmaglutinin or FLAG). Such hybrid polynucleotides produce fusion proteins that are useful, according to the invention, for improved expression and/or rapid isolation of a protein or protein fragment, encoded by the sequence of a gene. Hybrid polynucleotides are also useful as a source of antigen for the production of antibodies.

Nucleic acid constructs comprising a polynucleotide of genomic, cDNA, synthetic or semi-synthetic origin in association with a polynucleotide sequence encoding a foreign protein or a fragment thereof, (carrier sequence) can be generated by recombinant nucleic acid techniques well known in the art (See Ausubel et al., supra). According to this method, the cloned gene is introduced into an expression vector at a position located 3' to a carrier sequence coding for the amino terminus of a highly expressed protein, an entire functional moiety of a highly expressed protein or the entire protein. It is preferable to use a carner sequence from an *E. coli* gene or from any gene that is expressed at high levels in *E. coli*. It is often preferable to select a carrier sequence that will facilitate protein purification, either with antibodies, or with an affinity purification protocol that is specific for the carrier protein being used. For example, the purification protocol can be designed in accordance with the unique physical properties of the carrier protein (e.g. heat stability). Alternatively, the tag sequence may encode a protein (e.g. glutathione-S-transferase (GST)) which can be purified by either a chemical interaction (for example glutathione purification of GST). Alternatively, some carrier proteins, such as thioredoxin (Trx) can be selectively released from intact cells by osmotic shock or freeze/thaw procedures. Often, proteins that are fused to these carrier proteins can be purified away from intracellular contaminants by virtue of the physical attributes of the carrier protein (Ausubel et al., supra).

To ensure that a fusion protein is useful, according to the invention, it may be necessary to modify the expression protocol to produce a soluble protein. Due to the fact that high-level expression of certain proteins can lead to the formation of inclusion bodies, if a soluble protein is required it may be necessary to modify the following variables. The temperature at which expression is induced can affect inclusion body formation since inclusion body formation is induced at higher temperatures (37°C and 42°C) and inhibited at lower temperatures (30°C). In certain instances, lowering the total level of protein expression can lead to an increase in the proportion of soluble protein that is produced. The strain background of the cells in which the protein is being produced can affect the proportion of a particular protein that is expressed in a soluble form. Furthermore, the choice of carrier protein can affect the solubility of an expressed fusion protein (Ausubel et al., supra).

An additional problem that can be encountered when producing fusion proteins in *E. coli* is formation of an unstable protein, or a protein that is cleaved at the site of the junction between the carrier sequence and the sequence of the protein of interest. To decrease complications due to protein instability one can arrange for the fusion protein to be expressed as insoluble aggregates. Alternatively, one can express the fusion protein in *E. coli* strains that are deficient in proteases (Ausubel et al., supra).

Often it is useful to remove the carrier protein moiety from the protein of interest to facilitate biochemical and functional analyses. Methods for cleavage of fusion proteins to remove the carrier are known to those skilled in the art. The choice of a method is usually determined by the composition, sequence, and physical characteristics of the particular protein. Reagents such as cyanogen bromide, hydroxylamine or low pH can be used to chemically cleave fusion proteins. To avoid complications resulting from chemical cleavage (e.g. the presence of chemical cleavage sites in the protein of interest and/or the occurrence of side reactions resulting in protein modification), enzymatic cleavage methods can be used. Enzymatic cleavage protocols are advantageous because they can be carried out under relatively mild reaction conditions, and because they involve highly specific cleavage reactions. Enzymes useful for enzymatic cleavage of fusion proteins include factor Xa, thrombin, enterokinase, renin and collagenase (Ausubel et al., supra).

Recombinant constructs encoding fusion proteins wherein the carrier sequence is on the order of 9-15 codons, can be generated by PCR methods. According to this method, a PCR primer will be designed to contain at least 13 nucleotides that are identical to the target sequence on either side of the nucleotide sequence encoding the carrier sequence. Preferably, the PCR primer will also contain a restriction enzyme site to facilitate cloning of the amplified product into an appropriate expression vector. PCR will be carried out as described above and the sequence of the amplified product will be confirmed by sequence analysis as described in Section D entitled "Isolation of a Wild type Gene".

Alternatively, recombinant constructs encoding fusion proteins can be generated by site/oligonucleotide directed mutatagenesis (Ausubel et al., supra). According to the method of site directed mutatagenesis the DNA to be mutated is inserted into a plasmid which has an Fl origin of replication. A mutagenesis oligonucleotide is designed to contain 13 bp that are 100% identical to the target sequence, on either side of a sequence coding for the 9-15 codons of carrier sequence that is to be added by the mutatgenesis protocol.

A single stranded preparation of the vector is prepared by the following method. Following transformation of an appropriate bacterial strain (e.g. CJ236) with the recombinant plasmid and plating of the bacteria on LB agar plates, a single resulting colony is grown in 4x5 ml of LB plus ampicillin for

1 hour at 37°C with vigorous shaking. M13K07 helper phage (2 ml, approximately 10^{10}-10^{11} plaque forming units) is added and the bacteria are grown for an additional hour at 37°C with vigorous shaking. Following the addition of 7 ml of kanamycin (50 mg/ml), the bacteria are grown overnight at 37°C with vigorous shaking. The following day bacterial cultures are pooled and cells are separated by centrifugation. After the addition of 2.6 ml of 20% polyethylene glycol 200-800/2M NaCl to 20 ml of bacterial supernatant, the sample is incubated for 1 - 1.5 hours on ice. The sample is pelleted by centrifugation at 9000 rpm for 20 minutes. Following removal of the supernatant, residual supernatant are removed by centrifugation at 3000 rpm for 5 minutes. The pellet is resuspended in 400 ml of TE, extracted twice with phenol and four times with phenol:chloroform and ethanol precipitated. The resulting pellet is resuspended in 40 ml TE.

Mutagenesis is performed by using a muta-gene kit (Bio-Rad, Hercules, CA) according to the following method. To kinase the oligonucleotide primer, 1 m1 (200ng) of oligonucleotide is incubated in the presence of 2 m1 of 10 kinase buffer (0.5M Tris, pH 8.0, 70mM $MgCl_2$, 10mM DTT), 2 m1 10mM rATP, 2 m1 polynucleotide kinase and 13 ml H_20 for 37°C for 1 hour. To carry out the annealing and synthesis steps, 2.5 m1 of single-stranded template are mixed with 1 ml of kinased oligonucleotide, 1.0 m1 of 10X annealing buffer (200mM Tris-HCl, pH 7.4, 20 mM $MgCl_2$, 500mM NaCl) and 5.5 m1 H_20 for 10 min at 65°C. The reaction mixture is slow-cooled to 37°C. Once the sample has reached 37°C, the sample is spun briefly in a microfuge. Following the addition of 1.0 m1 of 10X synthesis buffer (5mM each dATP, dCTP, cGTP, dTTP, 10mM ATP, 100mM Tris-HCl, pH 7.4, 50 mM $MgCl_2$, 20mM DTT), 1.0 m1 T4 DNA ligase and 0.5 m1 of T4 DNA polymerase, the sample is incubated for 5 minutes on ice, 5 minutes at room temperature and 1 hour at 37°C. A 2 m1 aliquot of the sample is used to transform *E. coli.*

DNA is isolated from the transformed *E. coli* cells by mini prep methods known in the art (Ausubel et al., supra), and sequenced according to methods known in the art (described in Section D entitled "Isolation of a Wild Type Gene".

C. Production of a Nucleic acid Probe

The invention discloses nucleic acid probes. Preferably, the nucleic acid probes of the invention are specifically hybridizable to a mutant gene but not to a wild type form of a gene due to the presence of one or more polymorphisms. These allele specific probes can be used to screen DNA sequences of a gene which have been amplified by PCR, or are present in a genomic DNA or RNA test sample. Hybridization of a particular allele specific probe to an amplified gene sequence, under stringent conditions (described below), indicates that the polymorphism contained in the probe is

present in the amplified sequence. Hybridization of a particular allele specific probe to a test sample comprising genomic DNA or RNA, under stringent conditions (described below), indicates that the polymorphism contained in the probe, is present in the nucleic acid of the test sample. Nucleic acid probes that are specifically hybridizable to a wild type form of a gene but not to a mutant form of a gene are also useful according to the invention.

In another embodiment, the probes of the claimed invention will be specific for a nucleic acid region that is adjacent to a region that is thought to contain one or more polymorphisms. These probes will be useful for detecting the presence of one or more polymorphisms in the adjacent region by the method of primer extension (as described in Section F entitled "Identification and Characterization of Polymorphisms".

In other embodiments, probes of the claimed invention will be used to detect a gain or loss of a restriction enzyme site known to contain one or more polymorphisms of the claimed invention. Nucleic acid probes, according to this embodiment, are able to detect a restriction enzyme fragment that is of a size that can be easily separated on an agarose gel and visualized by Southern blot analysis. Probes that are useful according to this embodiment of the claimed invention can be specific for any region within a gene or outside of a gene.

The nucleic acids probes of the invention are useful for a variety of hybridization-based analyses including but not limited to Southern hybridization to genomic DNA, cDNA sequences or PCR amplification products, Northern hybridization to mRNA and RNase protection assays, DNA sequencing and isolation of genomic or cDNA clones of a gene. The probes may also be used to determine whether mRNA encoded for by a gene is present in a cell or tissue by the method of *in situ* hybridization. These techniques are well known in the art and can be performed as described in Ausubel et al., supra.

According to the methods of the above-referenced hybridization assays, polymorphisms associated with alleles of a gene, which either predispose to a particular disease (e.g. osteoarthritis) or are not associated with a particular disease (e.g. osteoarthritis), will be detected by the formation of a stable hybrid consisting of a polynucleotide probe comprising one or more polymorphisms and a target sequence, that also comprises one or more polymorphisms, under stringent to moderately stringent hybridization and wash conditions. If it is expected that the probes will be perfectly complementary to the target sequence, stringent conditions will be used. Hybridization stringency may be lessened if some mismatching is expected, for example, if variants are expected with the result that the probe will not be completely complementary. Conditions are chosen which rule out nonspecific/adventitious bindings, that is, which minimize noise. Since such indications identify neutral DNA polymorphisms as

well as mutations, these indications need further analysis (such as assays described in Section F entitled "Identification and Characterization of Polymorphisms") to demonstrate detection of a susceptibility allele of a gene.

Probes for alleles of a gene may be derived from genomic DNA or cDNA sequences from specific for the gene of interest. The probes may be of any suitable length, which span all or a portion of the region containing the gene. If the target sequence contains a sequence identical to that of the probe, the probes may be short, e.g., in the range of about 8-30 base pairs, since the hybrid will be relatively stable under even stringent conditions. If some degree of mismatch is expected with the probe, i.e., if it is suspected that the probe will hybridize to a variant region, a longer probe may be employed which hybridizes to the target sequence with the requisite specificity.

Probes according to the invention also include an isolated polynucleotide attached to a label or a reporter molecule which may be useful for isolating other polynucleotide sequences, having sequence similarity by standard methods, including but not limited to the above-referenced hybridization-based assays. Techniques for preparing and labeling probes (as described in Ausubel et al. Supra) are included below. A wide variety of labels and conjugation techniques are known by those skilled in the art and can be used in a various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting related sequences include oligolabeling, nick translation, end-labeling or PCR amplification using a labeled nucleotide. Alternatively, the protein-encoding sequence, or any portion of it, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3 or SP6 and labeled nucleotides.

A number of companies such as Pharmacia Biotech (Piscataway NJ), Promega (Madison WI) and US Biochemical Corp (Cleveland OH) supply commercial kits and protocols for these procedures. Suitable reporter molecules or labels include those radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles and the like. Patents teaching the use of such labels include US Patents 3,817,838; 3,350,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149 and 4,366,241. Also, recombinant immunoglobulins may be produced as shown in US Patent No. 4,816,567 incorporated herein by reference.

Probes comprising synthetic oligonucleotides or other polynucleotides of the present invention may be derived from naturally occurring or recombinant single- or double- stranded polynucleotides, or be chemically synthesized.

Portions of the polynucleotide sequence having at least approximately 5 nucleotides,

preferably 9-15 nucleotides, fewer than about 6 kb and usually fewer than about 1 kb, from a polynucleotide sequence encoding a gene are preferred as probes.

A DNA probe useful according to the present invention can be isolated from a gene or a polynucleotide constmct derived from a gene, or from a cDNA sequence specific for a gene or a cDNA construct specific for a gene by the methods of PCR or restriction enzyme digestion, as described above. Riboprobes useful according to the invention can be synthesized by the method of *in vitro* transcription, or by chemical synthesis methods, as described above.

An oligonucleotide probe useful according to the invention can be designed, as described above, and synthesized in a commercially available automated synthesizer.

Nucleic acid hybridization rate and stability will be affected by a variety of experimental parameters including salt concentration, temperature, the presence of organic solvents, the viscosity of the hybridization solution, the base composition of the probe, the length of the duplex, and the number of mismatches between the hybridizing nucleic acids (Ausubel et al., supra), and as described in Section A entitled "Design and Synthesis of Oligonucleotide Primers".

Southern blot analysis can be used to detect sequence variations in a gene from a PCR amplified product or from a total genomic DNA test sample via a non-PCR based assay. The method of Southern blot analysis is well known in the art (Ausubel et al., supra, Sambrook et al., 1989, Molecular Cloning. A Laboratory Manual., 2nd Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY). This technique involves the transfer of DNA fragments from an electrophoresis gel to a membrane support resulting in the immobilization of the DNA fragments. The resulting membrane carries a semipermanent reproduction of the banding pattern of the gel.

Southern blot analysis is performed according to the following method. Genomic DNA (5-20 mg) is digested with the appropriate restriction enzyme and separated on a 0.6-1.0% agarose gel in TAE buffer. The DNA is transferred to a commercially available nylon or nitrocellulose membrane (e.g. Hybond-N membrane, Amersham, Arlington Heights, IL) by methods well known in the art (Ausubel et al., supra, Sambrook et al., supra). Following transfer and UV cross linking, the membrane is hybridized with a radiolabeled probe in hybridization solution (e.g. under stringent conditions in 5X SSC, 5X Denhardt solution, 1% SDS) at 65^{0}C. Alternatively, high stringency hybridization can be performed at 68^{0}C or in a hybridization buffer containing a decreased concentration of salt, for example 0.1X SSC. The hybridization conditions can be varied as necessary according to the parameters described in Section A entitled "Design and Synthesis of Oligonucleotide Primers". Following hybridization, the membrane is washed at room temperature in 2X SSC/0.1% SDS and at 65^{0}C in 0.2X SSC/0.1% SDS, and exposed to film. The stringency of the wash buffers can also be

varied depending on the amount of the background signal (Ausubel et al., supra).

Detection of a nucleic acid probe-target nucleic acid hybrid will include the step of hybridizing a nucleic acid probe to the DNA target. This probe may be radioactively labeled or covalently linked to an enzyme such that the covalent linkage does not interfere with the specificity of the hybridization. A resulting hybrid can be detected with a labeled probe. Methods for radioactively labeling a probe include random oligonucleotide primed synthesis, nick translation or kinase reactions (see Ausubel et al., supra). Alternatively, a hybrid can be detected via non-isotopic methods. Non-isotopically labeled probes can be produced by the addition of biotin or digoxigenin, fluorescent groups, chemiluminescent groups (e.g. dioxetanes, particularly triggered dioxetanes), enzymes or antibodies. Typically, non-isotopic probes are detected by fluorescence or enzymatic methods. Detection of a radiolabeled probe-target nucleic acid complex can be accomplished by separating the complex from free probe and measuring the level of complex by autoradiography or scintillation counting. If the probe is covalently linked to an enzyme, the enzyme-probe-conjugate- target nucleic acid complex will be isolated away from the free probe enzyme conjugate and a substrate will be added for enzyme detection. Enzymatic activity will be observed as a change in color development or luminescent output resulting in a 10^3-10^6 increase in sensitivity. An example of the preparation and use of nucleic acid probe-enzyme conjugates as hybridization probes (wherein the enzyme is alkaline phosphatase) is described in (Jablonski et al., 1986, Nucleic Acids Res., 14:6115)

Two-step label amplification methodologies are known in the art. These assays are based on the principle that a small ligand (such as digoxigenin, biotin, or the like) is attached to a nucleic acid probe capable of specifically binding to a gene. Allele specific gene probes are also useful according to this method.

According to the method of two-step label amplification, the small ligand attached to the nucleic acid probe will be specifically recognized by an antibody-enzyme conjugate. For example, digoxigenin will be attached to the nucleic acid probe and hybridization will be detected by an antibody-alkaline phosphatase conjugate wherein the alkaline phosphatase reacts with a chemiluminescent substrate. For methods of preparing nucleic acid probe-small ligand conjugates, see (Martin et al., 1990, BioTechniques, 9:762). Alternatively, the small ligand will be recognized by a second ligand-enzyme conjugate that is capable of specifically complexing to the first ligand. A well known example of this manner of small ligand interaction is the biotin avidin interaction. Methods for labeling nucleic acid probes and their use in biotin-avidin based assays are described in Rigby et al., 1977, J. Mol. Biol., 113:237 and Nguyen et al., 1992, BioTechniques, 13:116).

Variations of the basic hybrid detection protocol are known in the art, and include

modifications that facilitate separation of the hybrids to be detected from extraneous materials and/or that employ the signal from the labeled moiety. A number of these modifications are reviewed in, e.g., Matthews & Kricka, 1988, Anal. Biochem., 169:1; Landegren et al., 1988, Science, 242:229; Mittlin, 1989, Clincal Chem. 35:1819; U.S. Pat. No. 4,868,105, and in EPO Publication No. 225,807.

D. Isolation of a Wild type gene

A wild type version of a candidate gene according to the invention can be isolated by cloning from an appropriately selected genomic library according to methods well known in the art. Methods of cloning are described in Section B entitled "Production of a Polynucleotide Sequence

The sequence of the cloned gene will be determined by sequencing methods well known in the art (see Ausubel et al., supra and Sambrook et al., supra). Methods of sequencing employ such enzymes as the Klenow fragment of DNA polymerase I, Sequenase® (US Biochemical Corp, Cleveland, OH), Taq polymerase (Perkin Elmer, Norwalk, CT), thermostable T7 polymerase (Amersham, Chicago, IL), or combinations of recombinant polymerases and proofreading exonucleases such as the ELONGASE Amplification System (Gibco BRL, Gaithersburg, MD). Preferably, the process is automated with machines such as the Hamilton Micro Lab 2200 (Hamilton, Reno NV), Peltier Thermal Cycler (PTC200; MJ Research, Watertown, MA) and the ABI 377 DNA sequencers (Perkin Elmer).

E. Isolation of a Mutant Gene

A mutant version of a candidate gene according to the invention can be isolated by cloning from an appropriately selected genomic library according to methods well known in the art. Methods of cloning are described in Section B entitled "Production of a Polynucleotide Sequence."

The sequence of the cloned gene will be determined by sequencing methods described in Section D entitled "Isolation of a Wild Type Gene."

F. Identification and Characterization of Polymorphisms

a. Identification of SNPs by in silico methods (isSNPs)

1. Identification of Polymorphisms in Candidate Genes

The starting point is a set of experimentally derived nucleic acid sequences. In order to be useful for SNP discovery by the invention, it is preferred that the sequences have complete chromatogram files from a gel or capillary electrophoresis sequencing machine. When this is not available, quality score data which assigns a score to each base in the sequence indicating the

likelihood of error for the basecall may be used. If neither of these data are available, the sequence may be used to assist the clustering of other sequences and in some cases to provide additional verification for a discovered SNP, but is not be used by the invention for the identification of the polymorphism.

The population of sequences used may constitute either a database of cDNA-derived sequences or genomic sequence. In a preferred embodiment, sequences used by the invention are from an assembled cDNA database, such as the LifeSeqGold database (Incyte Genomics, Inc(Incyte), Palo Alto, CA).

Derivation of Nucleic Acid Sequences

cDNA was isolated from libraries constructed using RNA derived from normal and diseased human tissues and cell lines. The human tissues and cell lines used for cDNA library construction were selected from a broad range of sources to provide a diverse population of cDNAs representative of gene transcription throughout the human body. Descriptions of the human tissues and cell lines used for cDNA library construction are provided in the LIFESEQ database (Incyte Pharmaceuticals, Inc. (Incyte), Palo Alto CA). Human tissues were broadly selected from, for example, cardiovascular, dermatologic, endocrine, gastrointestinal, hematopoietic/immune system, musculoskeletal, neural, reproductive, and urologic sources.

Cell lines used for cDNA library construction were derived from, for example, leukemic cells, teratocarcinomas, neuroepitheliomas, cervical carcinoma, lung fibroblasts, and endothelial cells. Such cell lines include, for example, THP-1, Jurkat, HUVEC, hNT2, WI38, HeLa, and other cell lines commonly used and available from public depositories (American Type Culture Collection, Manassas VA). Prior to mRNA isolation, cell lines were untreated, treated with a pharmaceutical agent such as 5'-aza-2'-deoxycytidine, treated with an activating agent such as lipopolysaccharide in the case of leukocytic cell lines, or, in the case of endothelial cell lines, subjected to shear stress.

Sequencing of the cDNAs

Methods for DNA sequencing are well known in the art. Conventional enzymatic methods employ the Klenow fragment of DNA polymerase I, SEQUENASE DNA polymerase (U.S. Biochemical Corporation, Cleveland OH), Taq polymerase (The Perkin-Elmer Corporation (Perkin-Elmer), Norwalk CT), thermostable T7 polymerase (Amersham Pharmacia Biotech, Inc. (Amersham Pharmacia Biotech), Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies Inc. (Life

Technologies), Gaithersburg MD), to extend the nucleic acid sequence from an oligonucleotide primer annealed to the DNA template of interest. Methods have been developed for the use of both single-stranded and double-stranded templates. Chain termination reaction products may be electrophoresed on urea-polyacrylamide gels and detected either by autoradiography (for radioisotope-labeled nucleotides) or by fluorescence (for fluorophore-labeled nucleotides). Automated methods for mechanized reaction preparation, sequencing, and analysis using fluorescence detection methods have been developed. Machines used to prepare cDNAs for sequencing can include the MICROLAB 2200 liquid transfer system (Hamilton Company (Hamilton), Reno NV), Peltier thermal cycler (PTC200; MJ Research, Inc. (MJ Research), Watertown MA), and ABI CATALYST 800 thermal cycler (Perkin-Elmer). Sequencing can be carried out using, for example, the ABI 373 or 377 (Perkin-Elmer) or MEGABACE 1000 (Molecular Dynamics, Inc. (Molecular Dynamics), Sunnyvale CA) DNA sequencing systems, or other automated and manual sequencing systems well known in the art.

The nucleotide sequences have been prepared by current, state-of-the-art, automated methods and, as such, may contain occasional sequencing errors or unidentified nucleotides. Such unidentified nucleotides are designated by an N. These infrequent unidentified bases do not represent a hindrance to practicing the invention for those skilled in the art. Several methods employing standard recombinant techniques may be used to correct errors and complete the missing sequence information. (See, e.g., those described in Ausubel, F.M. et al. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY; and Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY.)

Assembly of cDNA Sequences

Human polynucleotide sequences may be assembled using programs or algorithms well known in the art. Sequences to be assembled are related, wholly or in part, and may be derived from a single or many different transcripts. Assembly of the sequences can be performed using such programs as PHRAP (Phils Revised Assembly Program) and the GELVIEW fragment assembly system (GCG), or other methods known in the art.

Alternatively, cDNA sequences are used as "component" sequences that are assembled into "template" or "consensus" sequences as follows. Sequence chromatograms are processed, verified, and quality scores are obtained using PHRED. Raw sequences are edited using an editing pathway known as Block 1 (See, e.g., the LIFESEQ Assembled User Guide, Incyte Pharmaceuticals, Palo Alto, CA). A series of BLAST comparisons is performed and low-information segments and

repetitive elements (e.g., dinucleotide repeats, Alu repeats, etc.) are replaced by "n's", or masked, to prevent spurious matches. Mitochondrial and ribosomal RNA sequences are also removed. The processed sequences are then loaded into a relational database management system (RDMS) which assigns edited sequences to existing templates, if available. When additional sequences are added into the RDMS, a process is initiated which modifies existing templates or creates new templates from works in progress (i.e., nonfinal assembled sequences) containing queued sequences or the sequences themselves. After the new sequences have been assigned to templates, the templates can be merged into bins. If multiple templates exist in one bin, the bin can be split and the templates reannotated.

A resultant template sequence may contain either a partial or a full length open reading frame, or all or part of a genetic regulatory element. This variation is due in part to the fact that the full length cDNAs of many genes are several hundred, and sometimes several thousand, bases in length. With current technology, cDNAs comprising the coding regions of large genes cannot be cloned because of vector limitations, incomplete reverse transcription of the mRNA, or incomplete "second strand" synthesis. Template sequences may be extended to include additional contiguous sequences derived from the parent RNA transcript using a variety of methods known to those of skill in the art. Extension may thus be used to achieve the full length coding sequence of a gene.

Analysis of the cDNA Sequences

The cDNA sequences are analyzed using a variety of programs and algorithms which are well known in the art. (See, e.g., Ausubel, supra, Chapter 7.7; Meyers, R.A. (Ed.) (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853). These analyses comprise both reading frame determinations, e.g., based on triplet codon periodicity for particular organisms (Fickett, J.W. (1982) Nucleic Acids Res. 10:5303-5318); analyses of potential start and stop codons; and homology searches.

Computer programs known to those of skill in the art for performing computer-assisted searches for amino acid and nucleic acid sequence similarity, include, for example, Basic Local Alignment Search Tool (BLAST; Altschul, S.F. (1993) J. Mol. Evol. 36:290-300; Altschul, S.F.et al. (1990) J. Mol. Biol. 215:403-410.) BLAST is especially useful in determining exact matches and comparing two sequence fragments of arbitrary but equal lengths, whose alignment is locally maximal and for which the alignment score meets or exceeds a threshold or cutoff score set by the user (Karlin, S. et al. (1988) Proc. Natl. Acad. Sci. USA 85:841-845.) Using an appropriate search tool (e.g., BLAST or HMM), GenBank, SwissProt, BLOCKS, PFAM and other databases may be searched for sequences containing regions of homology to a query rbosm or RBOSM of the present

invention.

Other approaches to the identification, assembly, storage, and display of nucleotide and polypeptide sequences are provided in "Relational Database for Storing Biomolecule Information," U.S.S.N. 08/947,845, filed October 9, 1997; "Project-Based Full-Length Biomolecular Sequence Database," U.S.S.N. 08/811,758, filed March 6, 1997; and "Relational Database and System for Storing Information Relating to Biomolecular Sequences," U.S.S.N. 09/034,807, filed March 4, 1998, all of which are incorporated by reference herein in their entirety.

Protein hierarchies can be assigned to the putative encoded polypeptide based on, e.g., motif, BLAST, or biological analysis. Methods for assigning these hierarchies are described, for example, in "Database System Employing Protein Function Hierarchies for Viewing Biomolecular Sequence Data," U.S.S.N. 08/812,290, filed March 6, 1997, incorporated herein by reference.

Identification of Sequence Variants and Polymorphisms

The method comprise a series of filters to identify isSNPs from other sequencing variants and errors. The filters can be grouped into the following five sets of filters by the order of application in the method:

Preliminary Filters: the main filter in the first group removes the majority of base call errors by requiring a minimum phred quality score of 15. Additional filters at this stage deal with sequence alignment errors as well as errors resulting from improper trimming of vector sequence, chimeras and splice junctions.

Advanced Chromatogram Analysis: additional base call errors are then detected by examining the original chromatogram files in the vicinity of a putative SNP by an automated procedure resulting in a set of SNPs wherein the base call error rate is reduced to less than 5%.

Clone Error Filters: errors introduced during laboratory processing such as those caused by reverse transcriptase, polymerase or somatic mutation are among the most difficult to distinguish from true SNPs. The Clone Error filters use statistically generated algorithms to identify these sources of error. A small percentage of actual SNPs will be discarded at this stage.

Clustering Error Filters: these types of errors result from the incorrect clustering of close homologs, pseudo- genes or from contamination by nonhuman sequences. The filters developed to minimize these clustering errors are also statistically based. As above these filters may be reject a fraction of actual SNPs

Finishing Filters: these filters remove duplicate and redundant SNPs from the generated list of SNP, and remove SNPs which are from the hypervariable regions of hypervariable genes such as

immunoglobulin and T cell receptors.

Pre-processing steps

The sequences must first be trimmed to eliminate vector sequence, contamination and repetitive sequences. Then certain low information content sequences (for example, long runs of a single base, or two or three-base repeats) and repetitive sequences (for example Alu sequences in humans) must be massed (changed to N's) to prevent over-clustering errors. The clustering process then identifies the sets of sequences that are believed to be derived from the same original DNA sequence or gene. The sequences in each cluster or then aligned using a method such as phrap which also defines a consensus sequence. It will be well recognized by those skilled in the art that there are numerous existing programs for carrying out these processes, and the SNP discovery process described herein will work equally well with any of them. In the instant embodiment, the preferred processes are Blocked 1 for trimming and masking, a variety of different algorithms for clustering, and phrap for the alignment. It will be recognized by those skilled in the art that phrap and other alignment methods carry out a secondary clustering step which divides clusters into contigs, and carry out a secondary trimming step which defines the end points of the portion of each sequence which participates in the contig. The contigs then maybe searched for the occurrence of SNPs.

Errors in the trimming, clustering and alignment processes will cause SNP discovery errors, usually false positives (the prediction of SNPs where they do not exist). Additional filters which are the subject of the invention are designed to recognize and remove these errors by providing the ability to identify likely errors in the processes and to correct them.

In some instances, it is preferred, as an optional step, to unmask regions of sequences which were masked because of low information content or repetitive sequence) during the clustering process can be unmasked after clustering to allow discovery of SNPs within these regions.

Identification of Candidate SNP Sequences

The first step in identifying candidate SNP sequences is to redefine the end points of each sequence as the points within the previous end points where a stretch of at least 10 consecutive base calls, containing at least eight base changes, matches the consensus sequence exactly. Sequence trimming errors (both at single sequence stage and at the alignment stage contribute to the false positives when foreign sequence (vector, chimera or splice variant) is similar to the real sequence and the true boundary is difficult to determine. This step is a conservative approach to avoid false positives and also filters out lower-quality sequence that the ends. The reason the length of the match

with a consensus is measured in base changes is to avoid low significance matches on repetitive sequence such as polyA.

The next step is an each position of the alignment to compare the base calls of all the aligned sequences which are between their start and end positions and which have quality scores greater than a set threshold, and which have neighboring base calls which agree with a consensus sequence and where the neighboring base calls also have a quality score > the threshold. Preferably the threshold is a phred quality score greater than or equal to 15. The possibilities are A, C, G, T, and -(deletion).

The next step is a Clone Filter where if there has been more than one base call for a sequence position, then the clone for each sequence is identified in the sequences corresponding to each clone are compared. If the base calls for different sequences from the same clone disagree, then all the sequences for this clone at this base position are removed from consideration.

After all of these filters, positions for which there is more than one base call are candidate SNPs. The "wild type" base call is the one in the consensus sequence and the others are designated candidate SNPs. If the wild type base call is a deletion, then the SNP is considered to be an insertion at the previous base.

Automated Chromatogram Checking

The next filters require opening of the chromatogram files for the sequences identified as containing candidate SNPs. At each candidate SNP position, the chromatogram data of each sequence passing the Identification Filters is extracted. The first step in this process utilizes a program ABIdump to translate binary ABI chromatogram files into usable form.

Multiple Base Call Algorithm filter: the ABI base calls for each sequence are compared to the phred base calls. If the base calls do not agree at the SNP position and the two adjacent flanking positions, then the sequences are removed from consideration.

Intensity Filter: if the SNP is a single base change (this step is skipped for insertions and deletions), then the process intensity values for each of four bases at the call chromatogram location of the candidate SNP base are used to compute a ratio. If we call the intensity of wild type, "wt", the intensity of the SNP base "snp", the minimum of the other two "min", and the phred quality of the base call "Q", then the wild type sequences must have

(snp-min) < (wt-min)(Q-17)/37 and Q>=17 to be considered high-quality, and

(snp-min)<(wt-min)(Q-4)/37 and Q>=15 to be considered a low quality pass.

The basis for these formula is that if a base is mis- called, then there is likely to be a residual peak for the correct base. The larger the peak for the wild type base, the less likely that the call of the SNP is

correct. The actual thresholds in the formula are based on empirical data from clones which were sequence multiple times and which gave a set of confirmed SNPs and error rates for algorithm optimization.

The candidate SNP passes only if at least one wild type sequence passes and at least one SNP sequence passes. The quality of the candidate SNP is the lower of the highest wild type pass level and the highest SNP pass level (if there is a high-quality wild type sequence but only low quality SNP sequences, then the candidate is low quality. A SNP quality value is returned.

Clone Error Quality Filters (somatic mutation/reverse transcriptase/polymerase errors)

The purpose of these filters is to remove errors which are actually in the clone, that is, the clone sequence was correct but the clone does not represent the individual being sequenced. Three possible sources of these errors are somatic mutations, errors made by reverse transcriptase in the process of making cDNA, and DNA polymerase errors in those situations where the DNA has been amplified by PCR at some point prior to inserting in the cloning vector. Somatic mutations can be a particular problem in sequencing clones derived from cell lines.

Polymerase errors are specific to the type of sequencing protocol used. For example, reverse transcriptase is involved in EST sequencing but not genomic clone sequencing. Polymerase is involved in the creation of extension clones (polymerase is used in all sequencing reactions, but errors are less likely to arise because only a fraction of the templates are affected in contrast to the extension process where a single polymerase product becomes a template for the entire reaction). This filter is not applied to genomic sequences in the current embodiment on the premise that the genomic sequences do not have polymerase errors, and that somatic mutations are likely to have the same profile as real SNPs.

This filter also filters out rare SNPs as well as apparent SNPs which are not real. It is difficult to determine and confirm by experiments to what extent SNP candidates are too rare to be confirmed vs. simply not real. For many applications, very rare SNPs are of less utility than common ones such that this is not a problem; however in some applications it may be advisable to turn this filter off.

Base change sequence analysis filter

The premise of this filter is that probabilities of different mutations is different depending on the source. For example true SNPs may be mostly transitions whereas reverse transcriptase mutations could be primarily G to T mutations. While this does not allow one to determine for sure

that a given change is a true SNP, it allows one to evaluate the relative likelihood that a given mutation is a true SNP. SNP confirmation data suggest that G/T SNP candidates in which there is only one clone having the T allele have a very low probability of being real SNPs. The SNP candidates are excluded from the high confidence set (they are kept in a different file-their confirmation rate is well below 50 percent). The other set which had a very low confirmation rate is any A/T SNP.

Frequency Filter

This filter is based on the concept that true SNPs have a different frequency profile than clone errors and that a candidate SNP which is evident in only one clone in a deep alignment is less likely to be real than one which appears in one clone in a shallow alignment. The likelihood of finding a SNP at a given sequence location is a function of the number of chromosomes sequenced. This curve is distinctly non-linear as most SNPs are sufficiently frequent, to be found with relatively few sequences. The probability of an error of this type, however is essentially linear in the number of sequences since the chance of the change occurring in two different sequences is independent. This means that the probability that a candidate SNP observed in a single clone is a true SNP is lower if the alignment is deep then if a is shallow. Any SNP occurring in a single clone in an alignment of more than 20 clones (counting only high-quality sequences which have a chance of contributing a candidate SNP) is excluded from the high confidence set.

This filter is the basis of a secondary method used to develop the base change sequence analysis filter. Comparing the set of single clone SNPs from shallow alignment's with those from deep alignment's, which are more likely to be errors, will reveal base changes which are more likely to be associated with polymerase errors and somatic mutations.

Clustering Error Filters

These filters are intended to remove candidates SNPs which result from the incorrect clustering of similar sequences such as highly homogenous genes, similar genomic sequences, and contamination from other species where the sequences of the species have been mis- labeled as human.

Number of base change filter

This filter distinguishes homologous sequences from SNPs on the basis of the frequency of variants. True SNPs occur about one per kd when comparing to sequences or once per 2 kb if the length of sequences is included, and this fraction decreases as the depth of the alignment increases.

Since EST sequences tend to be about 500 bp or less in length, then it would be expected to have not more than one SNP per four sequences. The number of SNPs in the cluster is divided by the number of sequences in the cluster and SNPs for which this number is larger than one are discarded. The higher the number, the less likely the SNP is to be real. The threshold value of one was chosen because it appears to correspond to roughly a 50 percent success rate, however the threshold value could be adjusted to higher value to accept lower confidence SNPs.

Distance from next polymorphism filter

This filter calculates the number of SNPs for which the sequence is the only representative within a window of 100 bases on either side, and discards any of the SNPs for which there are more than one other SNP in this window. This threshold can be set higher, but the actual fraction of SNP candidates which are true SNPs drops off to less than 50 percent.

Haplotype clustering filter

When sequences from different sources are inappropriately clustered, it is possible to divide them into two or more clusters which are consistent. In particular, if we take any two differences between homologs and consider the haplotypes of the clones which overlap both SNPs, there are only two haplotypes. In other words, a 2x2 matrix of haplotypes is diagonal having only two non-zero entries. If there are only two sequences, then this is expected. For each SNP, a 2x2 haplotype matrix with each other SNP is computed. If it is diagonal, and there are more than two sequences, than the sum of the diagonal elements minus one is a "cluster total" for this SNP. This "cluster total" number has proven to be empirically correlated with the confirmation rate, probably because it predicts clusters which contain para-logs, homologs and contamination from other species. Candidates SNPs which have a cluster number of less than eight are kept. This threshold value for the cluster total can be varied.

Redundancy/finishing filters

Redundant SNP filter: SNPs in different contigs of the same gene which have the same base change and surrounding sequence are flagged as redundant. To accommodate possible splice variants this redundancy filter also applies to SNPs which have the surrounding sequence matches on only one side.

T cell receptor/immunoglobulin filters

Sequences containing SNPs are filtered to remove SNPs in sequences that are homologs to T cell receptors and immunoglobulin genes because both types of genes have hyper-variable regions which could result in false positives.

Output file

SNP related data: With each candidate SNP a variety of data is kept, including the number and sources of all contributing sequences (for example gene album, HTPS, FL, WashU/Merck, etc.), the surrounding sequence, measures of the ratio and quality scores for the "best" sequence representing each allele, etc.

Sequence related data: for each sequence associated with each SNP, the following data is kept including the distance in each direction to the end of the sequence, the distance in each direction to the next base different from the consensus and passing the initial quality filters, the library, tissue ID, donor ID and comments (for example tumor, diseases, normal).

These methods have been described in patent applications entitled "Method for the Identification of Sequence Polymorphisms using Polynucleotide Sequence Databases, and Single Nucleotide Polymorphisms Identified Thereby" (Attorney Docket Nos. GX-0006 P and GX-0010 P), and are hereby incorporated by reference.

b. Identification of polymorphisms in osteoarthritis associated genes by SSCP

The invention provides methods for detecting the presence of polymorphisms in candidate genes of the invention. The invention also provides methods for distinguishing polymorphisms which contribute to a particular disease (e.g. osteoarthritis) over polymorphisms which do not contribute to the disease.

1. Identification of Polymorphisms in Candidate Genes

Identification of polymorphisms in a candidate gene, according to the invention, will involve the steps of isolating the candidate gene, determining its genomic structure and identifying polymorphisms in the DNA sequences in any portion of the entire protein-coding region. The invention also provides methods for identifying polymorphisms in the DNA sequences corresponding to RNA splice junctions. The invention also provides methods for identifying polymorphisms in the DNA sequence corresponding to the regulatory (promoter) region of the candidate gene.

A candidate gene is isolated by cloning methods well known in the art (described above). Preferably the genomic structure of a candidate gene is determined by Southern blot analysis, as described in Section C. It is expected that the entire sequence of an open reading frame (ORF) of an average entire gene can be spanned by 16 PCR-amplified DNA fragments or amplimers of an average length of 225 bp. It is expected that a smaller gene can be spanned by 1-2 amplimers and that >50 amplimers are required to span extremely large genes. Primers useful for production of the amplimers of a particular candidate gene are designed based on preexisting knowledge of the sequence of the wild type gene, according to the primer design strategies described in Section A entitled "Design and Synthesis of Oligonucleotide Primers."

For PCR amplification of a region to be tested by SSCP it is preferable to design primers that amplify overlapping regions of the candidate gene. If a sequence variation is located in a region of a candidate gene that corresponds to the region to which the primers hybridize, the primers will likely not bind, the region containing this sequence variation will not be amplified and the variation will not be detected in PCR based assays. By producing overlapping amplimers it is expected that virtually all of the sequence variations in a particular candidate gene will be detected. The amount of overlap in the amplimers is somewhat variable (approximately 20%) and the precise location of the overlapping regions will depend on the location of regions comprising a sequence that is an appropriate primer sequence. It is a possibility that a polymorphism will be located at a position just adjacent to the primer site. Consequently, sequence information will be available for only 20 bp on one side of the polymorphism and for 104-279 bp on the other side of the polymorphism. However, this should be a sufficient amount of sequence information to allow definition of a unique sequence context in which to define the particular polymorphism.

Based on screening analysis of 92 samples (184 chromosomes), it is expected that about 50% of the amplimers will demonstrate polymorphisms, and that approximately 80% of these amplimers will detect changes at single positions while the remaining 20% will detect base changes at two positions. Based on these estimates, it is expected that there will be approximately 10 sequence variations per open reading frame. However, the number of amplimers that demonstrate polymorphisms with vary depending on the number of individuals tested, the ethnicity and structure of the population being tested, and the region of DNA being tested.

Preferably, each polymorphism will be detected in the context of an SSCP fragment. Polymorphism analysis by fluorescent SSCP (fSSCP, described in detail in Section F entitled "Identification and Characterization of Polymorphisms") uses PCR to generate an amplimer of DNA to be studied. The region to be tested is defined as the region between the primers (e.g. the region that

is incorporated into the PCR product and reflects the sequence of the DNA sample being tested). The PCR primers reflect the sequence of the DNA sample being tested and are incorporated into the PCR product as one end of each strand of DNA in the PCR product. If a polymorphism occurs in a primer binding site either the PCR primer does not bind due to the mismatch and the PCR will not produce a product, or the primer binds, an amplification step occurs wherein the primer is incorporated, but the amplified product does not contain the polymorphism which occurs at the primer binding site. Therefore, fSSCP provides a method of screening a DNA sequence located between PCR primers for the presence of polymorphisms.

The sensitivity of the technique of fSSCP for detecting a polymorphism is affected by length, such that there is a substantial decrease in the detection of polymorphisms in amplimers that are greater than 300 bp in length. However, different conditions for performing SCCP at high sensitivity with larger fragments, e.g. 800-1500 bp have also been described. If the length of DNA screened per amplimer is decreased then more amplimers are required to screen a region of a given size. Therefore, efficient screening of a gene dictates that the lower limit of the size of an amplimer is 125 bp. To attain specificity for a particular gene sequence, pnmers are usually 20-25 bp in length, and additional criteria such as G:C content, and intra- and inter-primer complementarity are important considerations in primer design (as described above). All of these considerations are addressed if the primer3 program (Copyright (c) 1996 Whitehead Institute for Biomedical Research) is employed to design pairs of primers suitable for use in a single PCR reaction. Typically, program parameters are set so that multiple amplimers are designed in the length range of 150-300bp, with predicted primer melting temperatures in the narrow range 60-62^{0}C. The narrow temperature range increases the likelihood that a single set of PCR conditions can be used to generate a wide variety of different amplimers.

If it is desirable to screen a contiguous stretch of DNA which is larger than the maximum fragment size desired for sensitive polymorphism detection by fSSCP (300 bp) it is necessary to use multiple amplimers (which are assayed separately) which span the region of interest. Since the primer sites in an amplimer are not tested, these sequences need to be contained within another amplimer. To test the primer sequence, overlapping amplimers are designed by an algorithm that evaluates a large number of amplimers generated by the primer3 program for the optimum overlapping set according to a cost function. Thus, a series of overlapping PCR amplification products can be used to test a contiguous stretch of DNA. Constraints on primer design are such that the absolute minimum overlap is rarely possible. As a result, some regions of overlap occur that results in 'double testing' of a particular segment of DNA. The detection efficiency is affected by the sequence context of the polymorphism; it is possible that a polymorphic site will be detected in only one of two different

amplimers which overlap the same site. One strategy that is useful for increasing polymorphism detection efficiency is to design overlapping amplimers to generate 2-fold coverage of all sequences.

SSCP does not detect 100% of polymorphisms. The invention provides for detection of polymorphisms with an efficiency of 95% under a single set of conditions using single coverage of sequences; a 2-fold screening strategy can be employed if it is necessary to increase this detection efficiency.

It is expected that the polymorphism can be located, and detected anywhere in the SSCP fragment except in the regions at each end that correspond to the sequence of the PCR primers. The precise location and identity of the sequence variation(s) of a particular SSCP fragment can be confirmed by sequencing the fragment as described in Section D entitled "Isolation of a Wild Type Gene". The sequence of a candidate gene will be compared to the known sequence of a wild-type version of the gene by using the following DNA/protein sequence analysis programs and methods.

There are a large number of freely available methods for performing sequence comparisons. These methods differ in their speed of execution, their sensitivity, and the type of comparisons they are able to make. For example one can compare two DNA sequences, two protein sequences, a DNA sequence to a protein sequence by conceptual translation, or DNA sequences as if they were protein sequences, again by conceptual translation. The BLAST suite of programs (Altschul et al., 1990, J.Mol.Biol. 215:403) are commonly used to perform the above-referenced type of analysis. Although the BLAST suite of programs provides a rapid method of determining multiple distinct similarities between two sequences, these programs are not guaranteed to find an optimal solution when comparing two sequences according to a particular set of parameters. PSI-BLAST is a more sensitive variant of BLAST that operates by iteratively searching the database while simultaneously refining the query pattern based on the results of the searches. Other packages of programs that are available and which have different specific properties include the HMMER, SAM, WISE, STADEN and FASTA packages, and the programs est_genome, dotter, e-PCR, Clustal, cross_match and phrap (Pearson, 1996, Methods Enzymol., 266:227).

If sequence information is available for the intron-exon boundaries and for a region of the intron (of approximately 30-150 bp) located immediately 5' of an intron-exon boundary, primers can be designed to produce amplimers useful for identifying polymorphisms located in the RNA splice junctions. Similarly, if the promoter region of a candidate gene has been sequenced, primers can be designed to produce amplimers useful for identifying polymorphisms located in the promoter region. Additional methods for detecting and isolating polymorphisms include, but are not limited to fluorescent polarization-TDI, mass spectroscopy denaturing gradient gel electrophoresis, chemical cleavage of

mismatch, constant denaturant capillary electrophoresis, RNase cleavage, heteroduplex analysis, sequencing by hybridization, DNA sequencing, representational difference analysis, and denaturing high performance liquid chromatography, described below in Section F entitled, "Identification and Characterization of Polymorphisms".

2. Methods of Determining if a Polymorphism Contributes to osteoarthritis

No two individuals (excluding identical twins or other clones) have the same sequence of DNA in their genome. Variability in gene sequences between individuals accounts for many of the obvious phenotypic differences (such as pigmentation of hair, skin, etc.) and many nonobvious ones (such as drug tolerance and disease susceptibility). In a population, the DNA sequence that occurs at the highest frequency at any given site is commonly referred to as the wild type sequence. The term "wild type sequence" can be misleading, however, because in different populations an alternative form of a DNA sequence may be predominant and thus considered wild type for that particular population. DNA polymorphisms are located throughout the genome, within and between genes, and the various forms may or may not result in differential gene function (as determined by comparing the function of two alternative forms of the same sequence). Most polymorphisms do not alter gene function and are called neutral polymorphisms. Some polymorphisms do have an effect on gene function, for example by changing the amino acid sequence of a protein, or by altering control sequences such as promoters or RNA splicing or degradation signals.

Polymorphisms can be used in genetic studies to identify a gene involved in a disease. If a polymorphism alters a gene function such that it increases disease susceptibility, then it will be present more often in individuals with the disease than in those without the disease. Alternatively, if a particular DNA variant is protective against a disease, it will be found more often in individuals without the disease than in those with the disease. Statistical methods are used to evaluate polymorphism frequencies found in diseased as compared to normal populations, and provide a means for establishing a causal link between a polymorphism and a phenotype. To detect a significant association between a disease and a polymorphic site, different tests may be used with either genotypic or allelic distributions. The simplest test consists of a t-test wherein the frequency of the polymorphic alleles in normal individuals and individuals with the disease phenotype is compared. A comparison of the genotypic distribution in normal individuals and individuals with the disease phenotype can also be performed using a chi-square test of homogeneity. These tests are implemented in all commercially or freely available statistical packages, for example SAS and S+, and are even included in Microsoft Excel. More sophisticated analyses will be performed by incorporating covariates such as linear regression or

logistic regression, and by accounting for the information provided by adjacent polymorphic sites (multipoint analysis). An example of this type of program is the freely available program "Analyze" by JD Terwilliger (currently available at the WWW site ftp://ftp.well.ox.ac.uk/pub/genetics/analyze).

If a polymorphism has a phenotypic effect, a bias will exist in the distribution of polymorphisms between groups that have and do not have the disease phenotype. This manner of analysis can be used to study a trait that is not necessarily a disease; any trait can be studied by comparing a group with a particular phenotypic form of a trait to a group with a different phenotypic form of that trait. It is important that the cases and controls are correctly matched with regards to ethnicity, environmental influences, and other factors which could effect the phenotype being studied. Studies which test polymorphism frequencies within groups exhibiting different phenotypes and use statistical methods to compare the group polymorphism frequencies and identify correlations with phenotypes, are known as "associations studies".

Some polymorphisms that occur in a single gene can alter the function of a gene sufficiently such that the polymorphism results in a disease (monogenic disease). However, many common human diseases are polygenic; that is they are the result of complex interactions of various forms of multiple genes. In the case of polygenic diseases, the alteration of a single gene may not be detrimental per se, but in combination with certain sequence variants of other genes, this altered DNA sequence may contribute to a disease phenotype. DNA variants leading to monogenic diseases are usually rare in a population due to the process of natural selection against those carrying the disease gene. As variants in genes that are involved in polygenic disease do not produce the disease phenotype unless they occur in the appropriate combination with other gene variants, normal individuals can carry a subset of the disease-contributing variants without suffering adverse effects. Thus, disease-contributing gene variants that are associated with polygenic diseases may exist at a high frequency in a normal population. Selection against these disease variant forms of a gene will only occur when they are present in the appropriate disease-causing combination and there may not necessarily be selection against these gene variants in individuals carrying a subset of the disease-contributing variants. Neutral DNA variants do not alter gene function or contribute to a disease, are under no selective pressure and occur at variable frequencies within populations.

Monogenic diseases tend to be rare within the population, and therefore few patients may be available for studies of these diseases. A polymorphism in a single specific gene is necessary and usually sufficient to cause a monogenic disease, such that associations between the variant gene and the phenotype are usually readily apparent. In cases where the expression of a mutation phenotype is complete, ("complete penetrance"), the polymorphism present in the disease gene will not be found

upon examination of a large number of normal individuals. If there is not complete penetrance then some apparently normal individuals will contain the mutation; the difference in frequency of occurrence of the variant gene in the disease group as compared to the normal population will reveal that the variant is associated with the disease.

In polygenic diseases, variation at different genes occurs in a combination which alters susceptibility to the disease. Although several genes may have variant forms which can contribute to a disease phenotype, it is not always necessary for a contributing variant to be present at every gene potentially contributing to the disease in a given affected individual. For example, a hypothetical disease could be caused by a particular combination of variants at three of four genes, designated as A, B, C, and D. Appropriate susceptibility variants in combination at any three of the genes can cause the susceptibility, i.e. one person with increased susceptibility may have susceptibility variants in genes A, B, and C, while another individual with increased susceptibility to the same disease will have susceptibility variants in genes B, C, and D. Therefore, although not all affected individuals will have the same susceptibility variants, the net result is that a diseased population will have susceptibility variant forms of genes A, B, C, and D at a higher frequency than an unaffected population (as detected by association studies).

Unlike monogenic diseases which result from polymorphisms that are not present in control populations, the polymorphisms which contribute to the polygenic disease are also present in a normal population. As described in the example above, an individual with susceptibility polymorphisms in only one or two of the genes potentially contributing to the disease susceptibility will be normal with regard to disease susceptibility. Therefore, normal populations can be used to identify polymorphic regions of the genome in the population, and these regions can then be specifically tested in larger patient and control populations. Typically, a gene is analyzed for the presence of polymorphisms by testing between 2 and 100 normal individuals in order to establish if a particular polymorphism is present for that gene in the population. Once a polymorphic site(s) has been defined, the polymorphic site is then tested in case (disease) and control (normal) populations and statistical analyses are performed to identify polymorphisms which occur at significantly different frequencies in the two populations.

The determination of the statistical significance of polymorphism frequency differences is dependent upon the size of the observed frequency difference between the populations, and on the size of the populations being studied. If a significant difference is found, then it can be concluded that an association exists between the polymorphism and the phenotype being studied. A statistically significant difference is a frequency difference at a particular site between populations which would be expected to occur by chance in only 5 out of 100 tests. That is, a difference which has a 95%

probability of being a true difference due to the affect of the gene.

The foregoing discussion describes a method of testing for an association between a polymorphism which is the direct contributor to a disease and the disease phenotype. However, polymorphisms which do not directly contribute to a disease can also be used to identify regions of the genome which contain genes that contribute to the disease by virtue of their proximity to disease-contributing polymorphisms.

In humans, DNA exists as 23 homologous pairs of linear molecules (chromosomes). Recombination is a process which results in reciprocal exchanges of short homologous DNA segments between these homologous DNA pairs. Only one of each of the 23 pairs of chromosomes is inherited by the offspring. The inherited chromosome is thus made up of tandemly arrayed segments of DNA derived from both of a pair of chromosomes. Consequently, DNA is transferred in segments from one generation to the next. Although the boundaries of each inherited segment may vary in each generation, the net effect is that sequences of DNA which are adjacent along the length of the molecule are inherited together at a higher frequency than sequences that are farther apart. If a region (continuous linear segment) of DNA has two or more polymorphisms that are close together, they will be co-inherited at a higher frequency than polymorphisms that are farther apart, as they are more likely to remain on the same segment of DNA during recombination. Therefore, if two or more polymorphisms are close together, they will occur together at a higher frequency in a population than would be expected by random segregation. This effect is known as linkage. Linkage studies are performed using multiply affected individuals within families; the most commonly used approach is to test markers located throughout the genome in many sets of affected sib pairs that share the same phenotype. Markers which are located in the region of a genome that contributes to the phenotype will be inherited in both siblings, along with the phenotype, at a higher frequency than expected by chance. Studies wherein data from many such families is compared can be used to implicate a region of a genome as one that contributes to a particular phenotype.

Linkage disequilibrium (LD) association studies provide another method for using polymorphisms in genetic studies. The method of LD involves making a correlation at the population level, between the alleles (alternative polymorphic forms of the same sequence site) present at different genomic sites. If site 1 has two variant forms, A and a, and site 2 has two variant forms B and b, the observation in a population that allele A at site 1 is more often found with allele B at locus 2 than with allele b is an example of LD. If allele B is a disease- contributing polymorphism, then testing at allele A may show an association with the disease.

Linkage disequilibrium may be generated in several ways. Maintenance of LD in a population

allows a disease association to be detected many generations after the formation of LD. The maintenance of LD is explained by linkage: the closer the two loci, the longer (in terms of number of generations) that particular LD is maintained. As a result, polymorphisms which do not directly contribute to a disease can be used to identify regions of the genome which contain a disease contributing polymorphism. If a polymorphism affects gene function such that it contributes to a phenotype being studied and is found to be associated with the phenotype, nearby (neutral) polymorphisms which are in LD with the disease polymorphism may also show an association with the disease. Conversely, if a polymorphism does not affect gene function but is found to be associated with a particular phenotype, this polymorphism is in LD with a different, but adjacent polymorphism that affects gene function such that it contributes to the phenotype being studied. If a neutral polymorphism is always inherited with a phenotype- contributing polymorphism, then the strength of the association of the neutral polymorphism to the phenotype will be equal to that of the polymorphism which affects gene function and is contributing to the phenotype. A polymorphism which shows an association with a phenotype (for instance with disease susceptibility) is a marker for that phenotype and implicates the region in which the polymorphism resides as a region containing a polymorphism which contributes to the phenotype. Additional flanking polymorphisms can be tested to determine the precise location of the true phenotype-contributing variant.

Linkage studies on families, and LD studies on populations have different degrees of resolution with regards to defining the size of a DNA region which contains the phenotype-contributing polymorphism. In general, linkage studies define an interval which potentially contains tens to hundreds of genes, while LD studies have been used to implicate single genes in the development of a particular phenotype.

3. Test Populations Useful for Polymorphism Genotyping

The invention provides methods of determining allelic frequencies by performing genotypic analyses in appropriate test populations.

Study cohorts:

Osteoarthritis Progression Cohort

Derived from a population of normal women aged 45-65. The original aims of the study, started in 1989, were to assess how many women around menopausal age would get arthritis and what factors predispose them to developing it. Also to look into factors that may be associated with progression of the disease.

A series of examinations, x-rays and questionnaires about lifestyle factors were carried out on 1003 women that were recruited to the study. This study has been going for 10 years. As a result, a unique, world-renowned and well respected study is available looking at the reasons why women develop osteoarthritis, potential risk factors and the genetics of the disease.

Prospective Severe Outcomes Cohort (case-control)

Five hundred joint replacement cases will be ascertained as will be age, ethnicity and gender matched controls. The clinical data envisaged are : HRT use, numbers of joints affected, occupation, injury history, age, BMI.

The list of studies relevant is shown in following table.

Study type	Population details	Reasonable objectives	Timing
Pilot[1]	100 progressors + 75 non-progressors, 100 normals, all female from the progression cohort. Detailed clinical data, 10 yr. follow-up: joint-space narrowing/yr., joints affected, BMD, fractures, CRP levels.	Large genetic effects for fast OA progression, proof of principle. Correlation with biomarkers. Possible novel target	6-8 Months
Biomarker study	800 women from progression cohort. DNA, serum, urine, 5 biomarkers	Correlation of genetics with biomarkers - v. useful for clinical trials.	12 Months
Progression hand & knee OA study	~800 women from progression cohort. Detailed clinical data, joint-space narrowing/yr., joints affected, BMD (hip and spine), fractures, CRP levels, full lipid measurements, incidence of fractures (assessed by X-rays), 10 yr. follow-up radiographs for all patients.	Genetic effects of OA progression. Risk of OA. Correlation with biomarkers. Possible novel target. Genetic effects of osteoporosis risk, correlation with BMD. Possibly genetic effects of lipid levels and CVD risk.	18 Months
Case-control	~500 cases (joint replacements) Vs 500 matched controls. Prospective study: DNA + 2 biomarkers. Clinical data required: steroid use, #joints affected,	Large genetic effects for OA risk, proof of principle. Possible novel target	~6-12 months for collection

occupation, injury history, age, BMI.

4. Assays Useful for Determining the Association of a Polymorphism with osteoarthritis

Clinical parameters

There is a general consensus that radiological changes are the preferred method for epidemiological studies on the basis of cross sectional and prospective correlations between severity of X-ray changes with the presence of pain and loss of function. In osteoarthritis, the loss of cartilage produces a narrowed space between bones. The pattern of joint space narrowing can help distinguish between osteoarthritis and rheumatoid arthritis. Bone spurs (osteophytes) also help diagnose osteoarthritis. Other relevant clinical end points are pain, disability, function, joint replacement and maintenance of joint structure. Stages of disease progression are as follows:

Early stage: focal swelling of articular cartilage followed by the appearance of irregularities in the surface.

Intermediate stage: progressive degradation and loss of articular cartilage. Also characterised by fibrillation (vertical splitting), detachment (horizontal splitting) and thinning of the cartilage.

Late stage: Articular cartilage is almost completely destroyed. Bony outgrowths (osteophytes) occur at the joint margins resulting in residual arthritis. Characterised by pain and limitation of joint movement.

Clinical measurements of OA

Quantitative traits of interest for the study of OA and its progression are:

- Osteophyte count.
- Joint space narrowing (mm/yr.)
  Number of joints affected
- Types of joints affected

In addition a series of biochemical markers can provide valuable information such as:

- COMP
- CRP
- HA
- Protocollagen Type II
- Bone resorption markers (e.g. collagen cross-links)

Confounding factors

Most currently recognised environmental risk factors for prevalent knee OA - obesity, knee injury, and physical activity, influence incidence more than radiographic progression. Furthermore, these factors might selectively influence osteophyte formation more than joint space narrowing. These findings are consistent with knee OA being initiated by joint injury, but with progression being a consequence of impaired intrinsic repair capacity.

Other known confounding factors are steroid (glucocorticoid) use and, in women, hormone replacement therapy. Glucocorticoids ameliorate erosion in animal OA models and suppress synthesis of matrix metalloproteinases (Saito et al. 1999). Estrogen replacement therapy, on the other hand, has been shown to have a moderate, but not statistically significant, protective effect against worsening of OA both in the Chingford (Hart et al. 1999) and Framingham (Zhang et al. 1998) studies.

5. Methods of Genotyping Polymorphisms

The invention discloses methods for performing polymorphism genotyping. These methods can be used to detect the presence of a polymorphism in a sample comprising DNA or RNA.

A DNA sample for analysis according to the invention may be prepared from any tissue or cell line, and preparative procedures are well-known in the art. The preparation of genomic DNA is performed as described in Section B.

RNA samples may also be useful for genotyping according to the invention. Isolation of RNA can be performed according to the following methods.

RNA is purified from mammalian tissue according to the following method. Following removal of the tissue of interest, pieces of tissue of ≤2g are cut and quick frozen in liquid nitrogen, to prevent degradation of RNA. Upon the addition of a volume of 20 ml tissue guanidinium solution per 2 g of tissue, tissue samples are ground in a tissuemizer with two or three 10-second bursts. To prepare tissue guanidiium solution (1 L) 590.8 g guanidinium isothiocyanate is dissolved in approximately 400

ml DEPC-treated H_2O. 25 ml of 2 M Tris-Cl, pH 7.5 (0.05 M final) and 20 ml Na_2EDTA (0.01 M final) is added, the solution is stirred overnight, the volume is adjusted to 950 ml, and 50 ml 2-ME is added.

Homogenized tissue samples are subjected to centrifugation for 10 min at 12,000 x g at 12°C. The resulting supernatant is incubated for 2 min at 65°C in the presence of 0.1 volume of 20% Sarkosyl, layered over 9 ml of a 5.7M CsCl solution (0.1g CsCl/ml), and separated by centrifugation overnight at 113,000 x g at 22°C. After careful removal of the supernatant, the tube is inverted and drained. The bottom of the tube (containing the RNA pellet) is placed in a 50 ml plastic tube and incubated overnight (or longer) at 4°C in the presence of 3 ml tissue resuspension buffer (5 mM EDTA, 0.5% (v/v) Sarkosyl, 5% (v/v) 2-ME) to allow complete resuspension of the RNA pellet. The resulting RNA solution is extracted sequentially with 25:24:1 phenol/chloroform/isoamyl alcohol, followed by 24:1 chloroform/isoamyl alcohol, precipitated by the addition of 3 M sodium acetate, pH 5.2, and 2.5 volumes of 100% ethanol, and resuspended in DEPC water (Chirgwin et al., 1979, Biochemistry, 18: 5294).

Alternatively, RNA is isolated from mammalian tissue according to the following single step protocol. The tissue of interest is prepared by homogenization in a glass teflon homogenizer in 1 ml denaturing solution (4M guanidiium thiosulfate, 25 mM sodium citrate, pH 7.0, 0.1 M 2-ME, 0.5% (w/v) N-laurylsarkosine) per 100mg tissue. Following transfer of the homogenate to a 5-ml polypropylene tube, 0.1 ml of 2 M sodium acetate, pH 4, 1 ml water-saturated phenol, and 0.2 ml of 49:1 chloroform/isoamyl alcohol are added sequentially. The sample is mixed after the addition of each component, and incubated for 15 min at 0-4°C after all components have been added. The sample is separated by centrifugation for 20 min at 10,000 x g, 4°C, precipitated by the addition of 1 ml of 100% isopropanol, incubated for 30 minutes at -20°C and pelleted by centrifugation for 10 minutes at 10,000 x g, 4°C. The resulting RNA pellet is dissolved in 0.3 ml denaturing solution, transferred to a microfuge tube, precipitated by the addition of 0.3 ml of 100% isopropanol for 30 minutes at -20°C, and centrifuged for 10 minutes at 10,000 x g at 4°C. The RNA pellet is washed in 70% ethanol, dried, and resuspended in 100-200 ml DEPC-treated water or DEPC-treated 0.5% SDS (Chomczynski and Sacchi, 1987, Anal. Biochem., 162: 156).

RNA prepared according to either of these methods can be used for genotyping by the methods of Northern blot analysis, S1 nuclease analysis and primer extension analysis (Ausubel et al., supra).

cDNA samples also may be prepared according to the invention, i.e., DNA that is complementary to RNA such as mRNA. The preparation of cDNA is well-known and well-

documented in the prior art.

cDNA is prepared according to the following method. Total cellular RNA is isolated (as described) and passed through a column of oligo(dT)-cellulose to isolate polyA RNA. The bound polyA mRNAs are eluted from the column with a low ionic strength buffer. To produce cDNA molecules, short deoxythymidine oligonucleotides (12-20 nucleotides) are hybridized to the polyA tails to be used as primers for reverse transcriptase, an enzyme that uses RNA as a template for DNA synthesis. Alternatively, mRNA species can be primed from many positions by using short oligonucleotide fragments comprising numerous sequences complementary to the mRNA of interest as primers for cDNA synthesis. The resultant RNA-DNA hybrid can be converted to a double stranded DNA molecule by a variety of enzymatic steps well-known in the art (Watson et al., 1992, Recombinant DNA, 2nd edition, Scientific American Books, New York).

Tissues or fluids which are useful for obtaining a DNA or RNA sample according to the invention include but are not limited to plasma, serum, spinal fluid, lymph fluid, external secretions of the skin, respiratory, intestinal and genitoruinary tracts, saliva, blood cells, tumors, organs, tissue and samples of *in vitro* cell culture constituents.

Genotyping methods which are useful according to the invention, i.e., for the detection of polymorphisms in nucleic acid samples isolated from individuals, are disclosed below.

Single Strand Conformation Polymorphism (SSCP) Screening and Fluorescent SSCP Screening (fSSCP)

SSCP Analysis

One technique for detecting DNA sequence variations in a biological sample is single strand conformation polymorphism (SSCP) (Glavac et al., 1993, Hum. Mut. 2:404; Sheffield et al., 1993, Genomics 16:325). SSCP is a simple and effective technique for the detection of single base changes. This technique is based on the principle that single-stranded DNA molecules assume specific sequence-based secondary structures (conformers) under nondenaturing conditions. The detection of point mutations by single stranded conformation polymorphism is believed to be due to an alteration in the structure of single stranded DNA. Molecules differing by only a single base substitution may assume different conformers and migrate differently in a nondenaturing polyacrylamide gel. Single stranded DNAs that contain sequence variations are identified by an abnormal mobility on polyacrylamide gels. SSCP detects all types of point mutations and short insertions or deletions that are located between the PCR primers (within the probe region) with apparently equal efficiency. This

technique has proven useful for detection of multiple mutations and polymorphisms, including SNPs. SSCP sensitivity varies dramatically with the size of the DNA fragment being analyzed. The optimal size fragment for sensitive detection by SSCP is approximately 125-300bp.

The mobility of a single stranded DNA or double stranded DNA fragment during electrophoresis through a gel matrix is dependent on its size. Small molecules migrate more rapidly than large molecules because they pass through the pores in the matrix more easily. Conventionally, electrophoresis of single stranded DNA involves a 'denaturing' gel which maintains the single strandedness of the molecules. The denaturant is typically urea in polyacrylamide gels, and typically formamide or sodium hydroxide in agarose gels. In contrast, according to the SSCP screening protocol, single-stranded DNA is analyzed on a 'nondenaturing' gel. When single stranded DNA is analyzed on a 'non-denaturing' gel, intramolecular interactions can occur. In particular, the single stranded DNA is able to (partially) bind to itself. Consequently, DNA that is separated by electrophoresis on an SSCP gel does not migrate as a linear molecule but rather, the mobility of the DNA on an SSCP gel is governed by both its size and tertiary structure (conformation). The tertiary structure of a single stranded DNA fragment is dependent on the sequence of the entire fragment. Therefore, if a polymorphism exists in a given fragment, the conformation will usually be altered. The technique is performed as follows.

One or more test DNA samples are prepared for analysis as described above, and subject to PCR amplification. Oligonucleotide primers are designed and synthesized as described above. Amplifications are performed in a total volume of 10 ml containing 50 mM KCl, 10 mM Tris-HCl, pH 9.0 (at 25^{0}C), 0.1 % Triton X-100, 1.5 mM $MgCl_2$, 0.2mM of dGTP, dATP, dTTP, 0.02 mM of non radioactive dCTP, 0.05 ml [a-^{33}P] dCTP (1,000-3,000 Ci mmol^{-1}; 10 mCi ml^{-1}), 0.2 uM each primer, 50 ng genomic DNA (or 1 ng of cloned DNA template) and 0.1 U Taq DNA polymerase. The PCR cycling profile is as follows : preheating to 94^{0}C for 3 min followed by 94^{0}C, 1 min; annealing temperature, 30 sec; 72^{0}C, 45 sec for 35 cycles and a final extension at 72^{0}C for 5 min. Annealing temperature is different for each PCR primer pair and can be optimized according to the parameters described above. Amplifications using Vent Taq polymerase (New England Biolabs) are performed in a total volume of 10 ul using the buffer provided by the manufacturer with 1 mM each of dGTP, dATP, dTTP, 0.02 mM dCTP, 0.25 ul [a-^{33}P] dCTP (1,000-3,000 Ci mmol^{-1};10 mCi ml^{-1}), 0.2 uM of each primer, 50 ng of genomic DNA (or 1 ng of cloned DNA template) and 0.1 U of Vent Taq DNA polymerase. Samples are heated to 98^{0}C for 5 min prior to addition of enzyme and nucleotides. The PCR cycling profile is 98^{0}C, 1 min; annealing temperature, 45 sec; 72^{0}C, 1 min for 35 cycles, followed by a final extension at 72^{0}C for 5 min. The length and temperature of each step of a PCR cycle, as

well as the number of cycles, is adjusted in accordance to the stringency requirements, as described above.

SSCP analysis is performed as follows. Ten ul of formamide dye (95% formamide, 20mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol) are added to 10 ul aliquots of radiolabeled PCR product. Following denaturation at 100^{0}C for 5 min, the reaction mixture is placed on ice. Two ul aliquots are loaded onto 8% acrylamide:bisacrylamide (37.5:1), 0.5X TBE (45 mM Tris-borate, 1 mM EDTA), 5% glycerol gels. Electrophoresis is carried out at 25W at 4^{0}C for 8 hours in 0.5X TBE. Dried gels are exposed to X-OMAT ARfilm (Kodak) and the autoradiographs are analyzed and scored for aberrant migration of bands (band shifts). SSCP may be optimized, as desired, as taught in Glavac et al., 1993, Hum. Mut. 2:404.

fSSCP Analysis

Techniques for screening multiple DNA samples simultaneously are also useful for performing rapid genotyping analysis on a large number of samples according to the invention. By pooling and multiplexing DNA samples in fluorescent SSCP (fSSCP) assays, the high throughput required for detecting sequence variations in a large number of samples is achieved (Makino et al., 1992, PCR Methods Appl. 2:10; Ellison et al., 1993, BioTechniques 15:684). According to the method of fSCCP, PCR products are visualized and analyzed using an ABI fluorescent DNA sequencing machine. Different primer pairs are identified by different color fluorochromes (4 different fluorochromes are now available). fSSCP offers the following advantages over SSCP. Unlike SSCP, fSSCP does not require handling of radioactive materials. Furthermore, the fSSCP technique allows for automated data and automated data analysis programs that detect aberrantly migrating samples. In contrast, SSCP evaluation involves visual examination by an individual, and does not provide a means for correcting for lane to lane variations in electrophoretic conditions, as does fSSCP analysis.

fSSCP Analysis is performed as follows.

Amplifications are performed in a total volume of 10 ul containing 50 mM KCl, 10mM Tris-HCl, pH 9.0 (at 25 ^{0}C), 0.1 % Triton X-100, 1.5 mM $MgCl_2$, 0.2mM of dGTP, dATP, dTTP, dCTP, 0.2 uM primer labeled with one of the fluorochromes HEX, FAM, TET or JOE, 50 ng genomic DNA (or 1 ng of cloned DNA template) and 0.1 U Taq DNA polymerase. The PCR cycling profile is as follows : preheating to 94^{0}C for 3 min followed by 94^{0}C, 1 min; annealing temperature, 30 sec; 72^{0}C, 45 sec for 35 cycles and a final extension at 72'C for 5 min. Annealing temperature is different for each PCR primer pair. Amplifications using Vent Taq polymerase (New England Biolabs) are performed in a total volume of 10 ul using the buffer provided by the manufacturer with 1 mM each of

dGTP, dATP, dTTP, dCTP, 0.2 uM primer labeled with one of the fluorochromes HEX, FAM, TET or JOE, 50 ng genomic DNA (or 1 ng of cloned DNA template) and 0.1 U of Vent Taq DNA polymerase. Samples are heated to 98°C for 5 min prior to addition of enzyme and nucleotides. The PCR cycling profile is 98°C, 1 min; annealing temperature, 45 sec; 72°C, 1 min for 35 cycles, followed by a final extension at 72°C for 5 min. Annealing temperature is different for each PCR primer pair. Two ul of fluorescent PCR products are added to 3 ul formamide dye (95% formamide, 20mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol), denatured at 100°C for 5 min, then placed on ice. Thereafter, 0.5-1 ml of Genescan™ 1500 size markers are added as an internal standard. Two ul of the mix is loaded onto 8% or 10% acrylamide:bisacrylamide (37.5:1), 0.5X TBE (45 mM Tris-borate, 1 mM EDTA), 5% glycerol gels and electrophoresis is performed on an ABI 377 DNA sequencing machine. Gel temperature is maintained between 4° and 10°C by an external cooling unit connected to the internal cooling plumbing and chambers. Electrophoresis is carried out at 2500-3500 volts for 4 - 10 hours in 0.5X TBE. Data is automatically collected and analyzed with Genescan and Genotype analysis software (ABI).

The fSSCP procedure identifies regions of 150-300 base pairs containing a sequence variation. To identify the exact sequence change, the fragment which demonstrates the aberrant migration is amplified again from the same biological sample, using non fluorescent primers. The sequence is then determined using standard DNA sequencing methods well known to those skilled in the art (Ausubel et al., supra).

Although SSCP and fSSCP techniques are preferred according to the invention, other methods for detecting sequence variations, including DNA sequencing, can be employed. Additional techniques for detecting DNA sequence variations useful according to the invention are described below.

Fluorescence Polarization-TDI

Fluorescence polarization-TDI is another preferred technique technique according to the invention for the detection of sequence variations. Template-directed primer extension is a dideoxy chain terminating DNA sequencing protocol designed to ascertain the nature of the one base immediately 3' to the sequencing primer that is annealed to the target DNA immediately upstream from the polymorphic site. In the presence of DNA polymerase and the appropriate dideoxyribonucleoside triphosphate (ddNTP), the primer is extended specifically by one base as dictated by the target DNA sequence at the polymorphic site. By determining which ddNTP is incorporated, the alleles present in the target DNA can be determined.

Fluorescence polarization is based on the observation that when a fluorescent molecule is exited by plane-polarized light, it emits polarized fluorescent light into a fixed plane if the molecules remain stationary between excitation and emission. However, because the molecule rotates and tumbles in solution, fluorescence polarization is not observed fully by an external detector. The fluorescence polarization of a molecule is proportional to the molecule's rotational relaxation time, which is related to the viscosity of the solvent, absolute temperature, molecular volume, and the gas constant. If the viscosity and temperature are held constant, then fluorescence polarization is directly proportional to the molecular volume, which is directly proportional to the molecular weight. If the fluorescent molecule is large (with high molecular weight), it rotates and tumbles more slowly in solution and flourescence polarization is preserved. If the molecule is small (with low molecular weight), it rotates and tumbles faster and fluorescence polarization is largely lost (depolarized).

In the FP-TDI assay, the sequencing primer is an unmodified primer wih its 3' end immediately upstream from a polymorphic or mutation site. When incubated in the presence of ddNTPs labled with different fluorophores, the allele-specific dye ddNTP is incorporated onto the TDI primer in the presence of DNA polymerase and target DNA. The genotype of the target DNA molecule can be determined simply by exciting the fluorescent dye in the reaction and determining whether a change in fluorescence polarization occurs. Chen et al., 1999, Genome Res., 9:492.

One or more test DNA samples are prepared for analysis as described above, and subject to PCR amplification. Oligonucleotide primers are designed and synthesized as described above. Amplifications are performed in a total volume of 10 ml containing 50 mM KCl, 10 mM Tris-HCl, pH 9.0 (at 25°C), 0.1 % Triton X-100, 1.5 mM $MgCl_2$, 0.2mM of dGTP, dATP, dTTP, 0.02 mM of non radioactive dCTP, 0.05 ml [a-^{33}P] dCTP (1,000-3,000 Ci mmol^{-1}; 10 mCi ml^{-1}), 0.2 uM each primer, 50 ng genomic DNA (or 1 ng of cloned DNA template) and 0.1 U Taq DNA polymerase. The PCR cycling profile is as follows : preheating to 94°C for 3 min followed by 94°C, 1 min; annealing temperature, 30 sec; 72°C, 45 sec for 35 cycles and a final extension at 72°C for 5 min. Annealing temperature is different for each PCR primer pair and can be optimized according to the parameters described above. Amplifications using Vent Taq polymerase (New England Biolabs) are performed in a total volume of 10 ul using the buffer provided by the manufacturer with 1 mM each of dGTP, dATP, dTTP, 0.02 mM dCTP, 0.25 ul [a-^{33}P] dCTP (1,000-3,000 Ci mmol^{-1};10 mCi ml^{-1}), 0.2 uM of each primer, 50 ng of genomic DNA (or 1 ng of cloned DNA template) and 0.1 U of Vent Taq DNA polymerase. Samples are heated to 98°C for 5 min prior to addition of enzyme and nucleotides. The PCR cycling profile is 98°C, 1 min; annealing temperature, 45 sec; 72°C, 1 min for 35 cycles, followed by a final extension at 72°C for 5 min. The length and temperature of each step of a PCR cycle, as

well as the number of cycles, is adjusted in accordance to the stringency requirements, as described above.

Following PCR amplification, unused PCR primers and dNTPs are destroyed by adding 2ml of PCR product to 2ml of SAP/Exonuclease cocktail (0.1U shimp alkaline phosphatase (1 U/ml,Amersham Pharmacia Biotech, Inc., Piscataway, NJ)and 0.2U *E. coli* exonuclease I (10 U/ml, Amersham)in SAP buffer (20mM TrisHCl, pH 8.0; 10 mM $MgCl_2$, Amersham))per well of a 384-well Black PCR plate (ABT). The mixtures are incubated at 37°C for 60 min before the enzymes are heat inactivated at 95°C for 15 min. The mixture is held at 4°C until used in the FP-TDI assay.

To the enzymatically treated PCR product, 2 ml of TDI reaction cocktail containing TDI buffer (50mM Tris-HCl (pH 9.0), 50mM KCl, 5 mM NaCl, 2 mM $MgCl_2$, 8% glycerol), 1 mM TDI primer, 12.5 nM of each of two allele specific dye-labled ddNTPs (ROX-ddGTP, BFL-ddATP, Tamra-ddCTP, or R6G-ddUTP; NEN Life Science Products, Inc., Boston, MA), and 0.32U Thermo Sequenase (Amersham). The reaction mixtures are incubated at 94oC for 15 min, followed by 34 cycles of 94°C for 30 seconds and 55°C for 15 seconds. Upon completion of the reaction cycles, the samples are held at 4°C.

After the primer extension reaction, 24 ml of TE buffer/methanol (2:1) is added to each sample well, and the fluorescence polarization is measured using a LJL Analyst (LJL Biosystems, Sunnyvale, CA).

<u>Denaturing Gradient Gel Electrophoresis</u>

Denaturing gradient gel electrophoresis (DGGE) is a gel system which allows electrophoretic separation of DNA fragments differing in sequence by a single base pair. The separation is based upon differences in the temperature of strand dissociation of the wild-type and mutant molecules. During electrophoresis, fragments migrating through the gel are exposed to an increasing concentration of denaturant in the gel. When the DNA fragments are exposed to a critical level of denaturant, the DNA strands begin to dissociate. This dissociation causes a significant reduction in the mobility of the fragment. The position in the gel at which the level of denaturant is critical for a particular DNA fragment is a function of the Tm of the DNA fragment and is therefore different for wild-type versus mutant fragments. Consequently, upon migration to the position at which the level of denaturant is at the critical point, for either the wild-type or the mutant fragment, the mobility of these two molecules will become different, thus resulting in their separation. The mutation detection rate of DGGE approaches 100%. Although the technique of DGGE is relatively simple to perform, and does not require radioisotopes or toxic chemicals, it does require some specialized equipment. Furthermore,

DGGE can only be used to analyze fragments between 100 and 800bp due to the resolution limit of polyacrylamide gels. DGGE is advantageous over other methods useful for detecting sequence variations because the behavior of DNA molecules on DGGE gels can be modeled by computer thereby making it possible to accurately predict the detectability of a mutation in a given fragment. Genomic DNA fragments can be efficiently transferred from the gel following DGGE as described in US Patent No. 5,190,856.

Chemical Cleavage of Mismatches

Chemical cleavage of mismatch (CCM) is another technique for detection of sequence variations that is useful according to the invention. CCM is based upon the ability of hydroxylamine and osmium tetroxide to react with the mismatch in a DNA heteroduplex and the ability of piperidine to cleave the heteroduplex at the point of mismatch. According to the method of CCM, sequence variations are detected by the appearance of fragments that are smaller than the untreated heteroduplex following denaturing polyacrylamide gel electrophoresis.

DNA fragments up to 1kb in size can be analyzed by CCM with a probable 100% detection rate for sequence variation. CCM is particularly useful for either detecting all of the sequence variations in a particular fragment of DNA or for determining that there are no sequence variations in a particular fragment of DNA.

Constant Denaturant Capillary Electrophoresis (CDCE) Analysis

CDCE analysis is particularly useful in high throughput screening, i.e., wherein large numbers of DNA samples are analyzed. CDCE analysis combines several elements of both replaceable linear polyacrylamide capillary electrophoresis and constant denaturant gel electrophoresis. The technique of CDCE is a rapid, high resolution procedure that demonstrates a high dynamic range, and is automatable. The method of CDCE, as described in detail in Khrapko et al., 1994, Nucleic Acids Res. 22:364, involves the use of a zone of constant temperature and a denaturant concentration in capillary electrophoresis. Linear polyacrylamide gel electrophoresis is performed at viscosity levels that permit facile replacement of the matrix after each run. For a typical 100 bp fragment of DNA, point mutation-containing heteroduplexes are separated from wild type homoduplexes in less than 30 minutes. Using laser- induced fluorescence to detect fluorescent-tagged DNA, the system has an absolute limit of detection of 3×10^4 molecules with a linear dynamic range of six orders of magnitude. The relative limit of detection is about 3/10,000, i.e., 100,000 mutant sequences are recognized among 3×10^8 wild type sequences. This approach is applicable to analysis of low frequency mutations, and

to genetic screening of pooled samples for detection of rare variants.

Rnase Cleavage

An additional method for genotyping that is useful according to the invention is RNase Cleavage. Various ribonuclease enzymes, including RNase A, RNase T1 and RNase T2 specifically digest single stranded RNA. When RNA is annealed to form double stranded RNA or an RNA/DNA duplex, it can no longer be digested with these enzymes. However, when a mismatch is present in the double stranded molecule, cleavage at the point of mismatch may occur.

RNase Cleavage is preferably performed with RNase A. Ribonuclease A specifically digests single stranded RNA but can also cleave heteroduplex molecules at the point of mismatch. The extent of cleavage at single base mismatches depends on both the type of mismatch, and the sequence of DNA flanking the mismatch. Sequence variations leading to mismatch are indicated by the presence of fragments that are smaller than the uncleaved heteroduplex on denaturing polyacrylamide gels.

According to the invention, RNase Cleavage involves forming a heteroduplex between a radiolabeled single stranded RNA probe (riboprobe) and a PCR product derived from a biological sample. If a point mutation is present in the PCR product, following treatment of the resulting RNA/DNA heteroduplex with RNase A, the RNA strand of the duplex may be cleaved. The sample is then denatured by heating and analyzed on a denaturing polyacrylamide gel. If the RNA probe has not been cleaved, it will be the same size as the PCR product. If the probe has been cleaved, it will be smaller than the PCR product. RNase Cleavage can be used to easily detect a 1 bp deletion. However, small insertions may not be as easily detected as small deletions, by RNASE Cleavage, as 'looping-out' occurs on the target strand rather than the probe strand.

Heteroduplex Analysis

Another method for genotyping according to the invention is heteroduplex analysis. Heteroduplex molecules, i.e., double stranded DNA molecules containing a mismatch, can be separated from homoduplex molecules on ordinary gels. The exact rate of detection of sequence variations by heteroduplex analysis is unknown, but is clearly significantly lower than 100%. Presumably, the sequence of DNA flanking the mismatch, rather than the actual mismatch affects the detectability. Mismatches that are located in the middle of a DNA fragment are detected most easily. Although heteroduplex analysis is less sensitive than some of the other genotyping methods described, it may be considered useful according to the invention due to its simplicity.

Mismatch Repair Detection (MRD)

Another technique that is useful for genotyping according to the invention is mismatch repair detection (MRD). MRD is an *in vivo* method that detects DNA sequence variation by the occurrence of a change in bacterial colony color. DNA fragments to be screened for variation are cloned into two MRD plasmids, and bacteria are transformed with heteroduplexes of these constructs. The resulting colonies are blue in the absence of a mismatch and white in the presence of a mismatch. MRD can be used to detect a single mismatch in a DNA fragment as large as 10 kb in size. MRD permits high-throughput screening of genetic mutations, and is described in detail in Faham et al., 1995, Genome Research 5:474.

Mismatch Recognition by DNA Repair Enzymes

Another technique that is useful for detecting sequence variations according to the invention is Mismatch Recognition by DNA Repair Enzymes. The E.coli mismatch correction systems are well-understood. Three of the proteins required for the methyl-directed DNA repair pathway: MutS, MutL and MutH are sufficient to recognize 7 of the possible 8 single base-pair mismatches (C/C mismatches are not recognized) and cut/nick the DNA at the nearest GATC sequence. The MutY protein, which is involved in a distinct repair system can also be used to detect A/G and A/C mismatches. Some mammalian enzymes are also useful for mismatch recognition: thymidine glycosylase can recognize all types of T mismatch and 'all-type endonuclease' or Topoisomerase I is capable of detecting all 8 mismatches, but does so with varying efficiencies, depending on both the type of mismatch and the neighboring sequence.

The MutS gene product is the methyl-directed repair protein which binds to the mismatch. Purified MutS protein has been used to detect mutations by several different methods. Gel mobility assays can be performed in which DNA bound to the MutS protein migrates more slowly through an acrylamide gel than free DNA. This method has been used to detect single base mismatches.

An alternative method for the use of MutS in mismatch recognition, which does not require gel electrophoresis, involves the immobilization of MutS protein on nitrocellulose membranes. Labeled heteroduplexed DNA is used to probe the membrane in a dot-blot format. When both DNA strands are used, all mismatches can be recognized by binding of the DNA to the protein attached to the membrane. Although C/C mismatches are not detected, the corresponding G/G mismatch derived from the other strand is recognized. This technique is particularly useful because it is simple, inexpensive, and amenable to automation. However, the detection efficiency of this method may be limited by the size of the DNA fragment. In particular, this method works well for very short

fragments.

Sequencing by Hybridization (SBH)

An alternative method for detecting sequence variations according to the invention is sequencing by hybridization (SBH). According to this method, arrays of short (8-10 base long) oligonucleotides are immobilized on a solid support in a manner similar to the reverse dot-blot protocol, and probed with a target DNA fragment. In particular, oligonucleotides are synthesized together and directly onto the support.

The synthesis system begins with a silicon chip coated with a nucleotide linked to a light-sensitive chemical group which is used to illuminate particular grid co-ordinates removing the blocking group at these positions. The chip is then exposed to the next photoprotected nucleotide, which polymerizes onto the exposed nucleotides.

In this manner, as a result of successive rounds of nucleotide additions, oligonucleotides of different sequences can be synthesized at different positions on the solid support. Thirty-two cycles of specific additions (i.e., 8 additions of each of the four nucleotides) should enable the production of all 65,536 possible 8-mer oligonucleotides at defined positions on the chip.

When the chip is probed with a DNA molecule, e.g., a fluorescently labeled PCR product, fully matched hybrids should give a high intensity of fluorescence and hybrids with one or more mismatches should give substantially less intense fluorescence. The combination of the position and intensity of the signals on the chip enables computers to derive the sequence of the DNA molecule being analyzed for the presence of sequence variations.

Allele-Specific Oligonucleotide Hybridization

The technique of allele-specific oligpnucleotide (ASO) hybridization or the 'dot-blot' is also useful for genotyping according to the invention. Under specific hybridization conditions, an oligonucleotide will only bind to a PCR product if the two are 100% identical. A single base pair mismatch is sufficient to prevent hybridization. A pair of oligonucleotides, one carrying the wild type base and the other carrying a single base change, as compared to the wild type sequence, can be used to determine if a PCR product is homozygous wild type, heterozygous or homozygous mutant for a particular base change. When performing conventional dot blots, the PCR product is fixed onto a nylon membrane and probed with a labeled oligonucleotide. When performing a 'reverse dot blot', an oligonucleotide is fixed to a membrane and probed with a labeled PCR product. The probe may be isotopically labeled, or non-isotopically labeled. The technique allows for the genotyping of multiple

PCR amplified samples for the presence of a single base change.

Allele-Specific PCR

Many methods for identifying sequence variations involve the analysis of PCR-amplified DNA. The allele-specific polymerase chain reaction (also called the amplification refractory mutation system or ARMS) comprises an assay that occurs during the PCR reaction itself. ARMS requires the use of sequence-specific PCR primers which differ from each other at their terminal 3' nucleotide and are designed to amplify only the normal allele in one reaction, and only the mutant allele in another reaction. When the 3' end of a specific primer is 100% identical to the target, amplification occurs. When the 3' end of a specific primer is not 100% identical to the target, amplification does not occur. Agarose gel electrophoresis is used to detect the presence of an amplified product. The genotype of a (heterozygous) wild-type sample is characterized by amplification products in both reactions, and a homozygous mutant sample generates product in only the mutant reaction.

This technique can be modified so that the 5' ends of the allele-specific primers are labeled with different fluorescent labels, and the 5' end of the common primers are biotin labeled. According to this alternate protocol, the wild-type specific and the mutant-specific reactions are performed in a single tube. The advantages of this approach are that a gel electrophoresis step is not required, and the method is amenable to automation.

Primer-Introduced Restriction Analysis

The method of primer-introduced restriction analysis (PIRA) can also be used for genotyping according to the invention. PIRA is a technique which allows known sequence variations to be detected by restriction digestion. By introducing a base change close to the position of a known sequence variation (for example by using a PCR primer containing a mismatch, as compared to the target sequence), it is possible to create a restriction endonuclease recognition site that indicates the presence of a particular sequence change. The combination of the altered base in the primer sequence and the altered base at the mutation site, creates a new restriction enzyme target site. This approach may be used to create a new restriction enzyme site in either the wild-type allele or the mutant allele. If a novel restriction enzyme site is introduced in the mutant allele then, following digestion with the appropriate restriction enzyme, the homozygous wild-type form would produce a single band of the full-length size, the homozygous mutant form would produce a single band of the reduced size and the heterozygous form would produce both full length and reduced sized bands. Band size will be analyzed by gel electrophoresis.

Oligonucleotide Ligation Assay

The technique of oligonucleotide ligation can also be used for genotyping according to the invention.

The method of oligonucleotide ligation is based on the following observations. If two oligonucleotides are annealed to a strand of DNA and are exactly juxtaposed, they can be joined by the enzyme DNA ligase. If there is a single base pair mismatch at the junction of the two oligonucleotides then ligation will not occur. According to the method of oligonucleotide ligation, the two oligonucleotides used in the assay are modified by the addition of two different labels. According to this method, the assay for a ligated product involves detecting a ligated product by assaying for the appearance of the labels of the two oligonucleotides on a single molecule rather than visualization of a new, larger sized DNA fragment by gel electrophoresis.

When ligation reactions are conducted in 96-well microtiter plates and ligation is scored by ELISA, the oligonucleotide ligation assay can be performed by a robot and the results can be analyzed by a plate reader and fed directly into a computer. This method is therefore extremely useful for detecting the presence of a sequence variation in a large number of samples. The oligonucleotide ligation assay is performed on PCR-amplified DNA. A modification of this assay, termed the ligase chain reaction, is performed on genomic DNA and involves amplification with a thermostable DNA ligase.

Direct DNA Sequencing

Genotyping according to the invention may also be carried out by directly sequencing the DNA sample in the region of the gene of interest, using DNA sequencing procedures well-known in the art (described above in Section D, entitled "Isolation of a Wild Type Gene").

Mini-Seqnencing

The technique of mini-sequencing (also known as single nucleotide primer extension) can also be used to detect any known point mutation, deletion or insertion, according to the invention. Obtaining sequence information for just a single base pair only requires the sequencing of that particular base. This can be done by including only one base in the sequencing reaction rather than all four. When this base is labeled and complementary to the first base immediately 3' to the primer (on the target strand), the label will not be incorporated. Thus, a given base pair can be sequenced on the basis of label incorporation or failure of incorporation without the need for electrophoretic size separation.

5' Nuclease Assay

Genotyping according to the invention can also be performed by the method of 5' nuclease assay. The 5' nuclease assay is a technique that monitors the extent of amplification in a PCR reaction on the basis of the degree of fluorescence in the reaction mix. A low level of fluorescence indicates no amplification or very poor amplification and a high level of fluorescence indicates good amplification. This system can be adapted to permit identification of known sequence variations, without the need for any post-PCR analysis other than fluorescence emission analysis.

PCR amplification is detected by measuring the 5' to 3' exonuclease activity of Taq polymerase. Taq polymerase cleaves 5' terminal nucleotides of double stranded DNA. The preferred substrate for Taq polymerase is a partially double stranded molecule. Taq polymerase cleaves the strand that contains the closest free 5' end. According to the 5' nuclease assay, an oligonucleotide 'probe' which is phosphorylated at its 3' end so as to render it incapable of serving as a DNA synthesis primer, is included in the PCR reaction. The probe is designed to anneal to a position between the two amplification primers. When an actively extending Taq polymerase molecule reaches the probe molecule, it partially displaces the probe and then cleaves the probe at or near the single stranded/double stranded cleavage site until the entire probe is broken up and removed from the template. The polymerase continues this process of displacement and cleavage until the entire probe is broken up and removed from the template. The probe is labeled in a manner that permits detection of the removal of the probe. In particular, the probe is labeled at different positions with two different fluorescent labels. One label has a localized quenching effect on the fluorescence of the other (reporter) label. This effect is mediated by energy transfer from one dye to the other, and requires that the two dyes are in close proximity to each other. If the probe is cleaved at a position between the reporter and the quencher dyes, the two dyes become physically separated thereby resulting in an increase in fluorescence which is proportional to the yield of the PCR product.

Representational Difference Analysis (RDA)

Genotyping according to the invention can also be carried out by Representational Difference Analysis (RDA). RDA is described in detail in Lisitsyn et al., 1993, Science 259:946, and an adaptation which combines selective breeding with RDA is described in Lisitsyn et al., 1993, Nature Genet. 6:57. RDA identifies sequence dissimilarities through the application of a powerful approach to subtractive hybridization. According to the method of RDA, one first creates simplified representations, called amplicons, from two samples that are being compared. An amplicon can comprise, for example, the set of BglII fragments that are small enough to be amplified by the PCR.

The iterative subtraction step begins with the ligation of a special adaptor to the 5' end of fragments contained in the amplicon derived from the test sample (tester amplicon). The tester amplicon is then melted and briefly reannealed in the presence of a large excess of amplicon, derived from the wild type sample (driver amplicon). Those tester fragments that reanneal (presumably fragments absent from the wild type, driver amplicon) can serve as a template for the addition of the adaptor sequence to the 3'-end of the "partner" fragment. As a result, these tester fragments can be exponentially amplified by PCR. This procedure is then repeated to achieve successively higher enrichment.

RDA may be used to clone sequences that are either wholly absent from the wild type sample or are present in the wild type DNA, but are contained in a restriction fragment that is too large to be amplified in the amplicon. The former case may arise from a total deletion; the latter from a restriction fragment length polymorphism with the short allele present in the tester but not the wild type DNA. RDA is useful for subtracting DNA from an individual with a particular disease from normal DNA so as to identify regions showing homozygous or heterozygous deletions; locating fragments present in a parent with a dominant disorder but absent in his unaffected offspring; and locating mRNAs expressed in normal tissue but not present in tissue isolated from an individual with a particular disease.

Denaturing High Performance Liquid Chromatography

According to the scanning method of Denaturing High Performance Liquid Chromatography (DHPLC), partial heat denaturation and a linear acetonitrile column are used to identify polymorphisms in DNA fragments. DHPLC provides a method of comparative DNA sequencing based on the capability of ion-pair reverse phase liquid chromatography on alkylated nonporous poly(styrene divinylbenzene) particles to resolve homo- from heteroduplex molecules under conditions of partial denaturation. This method can potentially be automated to allow for rapid analysis of a large number of samples (Underhill et al., 1996, Proc. Natl. Acad. Sci. USA, 93:196).

Mass Spectroscopy

Matrix-assisted laser desorption-ionization-time-of-flight (MALDI-TOF) mass spectroscopy is another method according to the invention by which genotyping can be performed. The method of MALDI-TOF mass spectroscopy is based on the irradiation of crystals formed by suitable small organic molecules (referred to as the matrix) with a short laser pulse at a wavelenght close to the resonant adsorption band of the matrix molecules. This causes an energy transfer and desorption process producing matrix ions. Low concentrations of nucleic acid molecules are added to the matrix molecules while in solution and become embedded in the solid matrix crystals upon drying of the

mixture. The intact nucleic acids are then desorbed into the gas phase and ionized upon irradiation with a laser allowing their mass analysis. MALDI is used primarily with time-of-flight spectrometers where the time of flight is related to the mass-to-charge ratio of the nucleic acids molecules. Reviewed in Griffin T.J. and Smith L.M., 2000, Trends Biotech 18:77.

Genotyping can be performed by any of the following MALDI-TOF mass spectroscopy approaches including sequencing of PCR products (Fu, D-J et al., 1998, Nat. Biotechnol. 16:381; Kirpekar, F. et al., Nucleic Acids Res. 26:2554), direct mass-analysis of PCR products (Ross, P.L. et al., 1998, Anal. Chem. 70:2067), analysis of allele-specific PCR (Taranenko, N.I. et al, 1996, Genet. Anal. Biomol. Eng. 13:87) or LCR (ligase chain reaction; Jurinke, C. et al., 1996, Anal. Biochem. 237:174) products, analysis of RFLP-PCR products (Srinivasan, J.R. et al., 1998, Rapid Commun. Mass Spectrom. 12:1045), minisequencing (Haff, L.A. and Smirnov, I.P., 1997, Genome Res. 7:378; Higgens, G.S. et al., 1997, BioTechniques 23:710), analysis of PNA (peptide nucleic acid) hybridization probes (Griffin, T.J. et al., 1997, Nat. Biotech. 15:1368; Ross, P.L., Anal. Chem. 69:4197; Jiang-Baucom, P. et al., 1997, Anal. Chem. 69:4894), or direct analysis of invasive cleavage products (Griffin, T.J. et al, 1999, Proc. Natl. Acad. Sci. USA 96:6301).

6. Methods of Specifying a Polymorphism

The invention provides methods for specifying a particular polymorphism. By "specifying an polymorphism" is meant defining a polymorphism in the context of a larger region of nucleic acid which contains the polymorphism, and is of sufficient length to be easily differentiated from any other position in the genome.

A unique nucleotide position (e.g. a polymorphic site) in the human genome can be specified by describing a unique sequence of DNA within the genome, and providing the location of the unique nucleotide position relative to that sequence. Preferably this is done by providing the sequence identity of a length of unique DNA containing the polymorphism, and indicating which of the nucleotide sites is polymorphic.

A calculation can be made to determine a sequence length which will be unique in the 3 billion nucleotide human genome. If it is assumed that the genome contains equal numbers of the nucleotides A, G, C and T, and that they occur randomly in the genome, one can determine the probability of any given sequence of a defined length occurring in the genome; a random 12mer will appear in a random 3,000,000,000 bp genome 179 times, a random 15 mer will appear in a random 3,000,000,000 bp genome 3 times and a random 16mer will appear in a random 3,000,000,000 bp genome 1 time.

Thus, it would appear that specifying 16 bp would uniquely define a sequence in the genome.

However, the genome is not composed of random sequence and does not contain equal amounts of A, G, C and T. In fact, 10-12 bp sequences are likely to be specific for 95% of genes. Some sequences may even be specified by as few as 8 nucleotides. The minimum sequence length that is useful according to the invention for identifying polymorphisms in most gene and intergenic sequences is approximately 9-15 bp.

In the case of repeat sequences and sequences associated with gene families, the probability of observing a particular sequence is greatly increased and it becomes difficult to specify a polymorphism in the context of a sequence that is only on the order of 9-15 bp. There are many types of repeats including tandem repeats, where a larger sequence block has within it smaller repeat units (e.g. microsatellites). Tandem repeats usually occur within non-genic areas, but can also occur within genes and subsequently affect gene function; they can be 10-1000s of bp long, or, if located in centromeres and telomeres, be megabase sized. Some repeats are composed of blocks which do not have sub-repeat units and are non-functional (e.g. ~300 bp Alu repeats). These occur by duplication/dispersal throughout the genome.

It may be difficult to specify a polymorphism that occurs in a gene that is a member of a gene family. Through the mechanism of gene duplication, gene families, comprising multiple copies of a gene in which some, but not all of the DNA sequence has diverged, have been formed. Thus, certain regions of a gene may be conserved in different gene family members. With time, a duplicated gene can lose function and the sequence of the duplicated gene can deteriorate; the amount of homology between the original gene and the duplicated version depends upon the time since duplication. Other duplications maintain function and retain some level of similarity with the original gene in the important domains. Some related genes can share nearly 100% homology across a region that is hundreds of bp long, and yet have no significant homology at any other location. In these cases, it may be necessary to specify dozens or more nucleotides to provide a unique sequence.

To identify a unique sequence, a search must be done wherein a specific sequence is compared to all known human sequences and the minimum unique sequence is defined. However, in the absence of a complete sequence for the human genome, it cannot be guaranteed that a sequence is truly unique. Empirical experimentation can be used to determine the minimum sequence for specificity/uniqueness. In the case of a gene family member, if sequence information is available for the region corresponding to the region of interest in other members of the gene family, than it may be possible to define a unique short (9-15 bp) sequence that contains a polymorphism and has specificity. In the event that a particular region cannot be defined as unique, a larger region of nucleic acid which contains the polymorphism will be required to define a polymorphism in a gene that is a member of a

gene family. It is predicted that a sequence of 9-15 bp will be sufficient to define a polymorphism in 99% of all cases.

Methods of specifying a polymorphism that involve using sequences which either encompass or overlap the polymorphic site to be tested or do not encompass or overlap the polymorphic site to be tested are useful according to the invention and are described below.

Oligonucleotide Hybridization.

An oligonucleotide is designed such that it is specific for a target sequence, and hybridizes only at the target sequence site. This oligonucleotide will not hybridize if the target sequence differs at the position in the sequence to be tested. Another oligonucleotide is designed such that it hybridizes with the polymorphic form of the sequence. A DNA sample is tested for hybridization with each of the two probes independently. If the DNA hybridizes to only one of the probes, it can be concluded that the individual is homozygous for the corresponding sequence. If both probes hybridize to a test DNA sample, then the individual is heterozygous. Hybridization will be detected by the method of Southern blot analysis (as described in Section C entitled "Production of a Nucleic Acid Probe").

Specifying a Polymorphism by PCR

An alternative method for specifying a particular polymorphism involves a PCR-based strategy. According to this method, a region of a candidate gene to be tested is amplified by PCR (as described). The amplified fragment is digested with a restriction enzyme that will not cut a fragment that contains a polymorphism, due to the location of the polymorphism within the recognition site of this restriction enzyme. The products of the digestion reaction mixture are size separated in an agarose gel, stained with ethidium bromide, and visualized under ultraviolet light to determine if the amplified product has been digested. According to this method, the PCR primers provide the specificity for a particular polymorphism by virtue of the specific sequence of the two primers, as well as by the location of the primer binding sites in the target DNA. Although, multiple sites for primer binding may exist in a target DNA sequence, only the sites that are close enough together will produce an amplified product that includes the nucleic acid region containing the polymorphism.

Alternatively, a PCR reaction is carried out with PCR primers that contain polymorphisms. According to this embodiment, if the template nucleic acid lacks the polymorphism present in the primers there will be no PCR product. Thus, according to this embodiment of the invention, the absence of a PCR product indicates that a polymorphism is not present in the target sequence.

Primer Extension

A DNA fragment comprising the region containing a polymorphism is PCR amplified from an individual to be tested. The PCR product is denatured and one strand is retained for analysis. An oligonucleotide probe is designed such that it is specific for a region in the sequence and hybridizes such that its 3' terminal nucleotide is paired with the nucleotide adjacent to the one to be tested. The PCR product and probe are combined with a polymerase and terminating, differentially colored, nucleotides. The polymerase extends the probe by one base, and only the base which is complementary to the site being tested is added. The reaction is washed, and the color of the reaction indicates the nucleotide that has been added and the sequence at the position of interest.

The PCR step provides one level of specificity by amplifying a region (1 - 10000 bp as desired between the PCR primers) from a complex (3,000,000,000 bp) mixture. The PCR probes primers must be unique in both their hybridization specificity and their proximity to one another. Since proximity of the two PCR primers is needed (i.e. a distance across which a polymerase can extend to join the primers), shorter PCR primers can be used, e.g. in theory a small enough region could be amplified with a 8-10 bp binding site for a PCR primer. To ensure that a primer hybridizes with specificity, a primer must be at least 5 bp.

A second level of specificity is provided by the primer which is extended in the primer extension reaction. Since this primer is hybridizing to a short piece of DNA, it can be short and unique for the fragment with which it binds. The primer is at least 5bp and preferably 8bp. Although the primer used for the primer extension step is located probe adjacent to the polymorphic site, the PCR primers should not overlap with the polymorphic site being tested.

Southern Blotting

One method for detecting a previously defined polymorphism involves Southern blot analysis of wild type and mutant DNA following digestion with a restriction enzyme which has a recognition sequence which includes the polymorphic site to be tested. According to this method, a particular restriction enzyme cuts wild type DNA but does not cut mutant DNA due to the presence of a polymorphism within the recognition site of this restriction enzyme. Many restriction enzymes exist which recognize 4bps. The resulting fragments will be size separated in an agarose gel, transferred to a membrane and probed with a nucleic acid probe. If the site is uncut, the fragment is one length and if the site is cut the fragment will be of a shorter length.

The nucleic acid hybridization probe will provide specificity to the particular polymorphism being tested by defining the polymorphism in the context of a larger stretch of nucleic acid sequence.

The nucleic acid probe may comprise the nucleic acid sequence corresponding to the region known to contain the polymorphism. The sequence-specific probe may be located 10, 100, 1000, or even 100s of thousands of bases from the region containing the polymorphism. If the probe is located some distance from the region containing the polymorphism, an intervening recognition site for the restriction enzyme cannot be located between the probe hybridization site and the region of interest containing the polymorphism site. Typically, a hybridization probe useful according to this method will be much larger than the minimum length of a sequence (9-15 bp) required to give specificity to, or define a particular polymorphism.

Alternatively, a chemical or enzyme which recognizes a unique pair of nucleotides at the site of a polymorphism, can be used to detect the polymorphism. According to this method, the amount of sequence required for recognition by a chemical or enzyme is 2 bp (providing that the 2 bp sequence is unique in a region large enough to produce a fragment which can then be bound by a specific probe).

According to a variation of the above method, a labeled chemical or enzyme which binds to one sequence of the polymorphic recognition site and not another is used. This method involves the steps of digesting the DNA with a restriction enzyme, and adding a labeled, sequence-specific binding protein (e.g. a restriction enzyme that lacks cleavage capability). The sequence-specific binding protein will bind to multiple sites in the genome, including the site to be tested. The fragments will be separated on a gel and then probed with a probe specific for the test sequence. If the fragment identified by the second probe is identical to a fragment identified by the first probe (e.g. the labeled chemical or enzyme), then the sequence being tested for is present.

7. Determination of the Phenotypic Outcome of a Polymorphism

To determine the phenotypic outcome of a polymorphism according to the invention, it is necessary to screen suitable populations to obtain a statistically significant measure of the association of a polymorphism with a particular disease (e.g osteoarthritis). The invention provides methods for performing polymorphism genotyping in appropriate populations (described above). The invention also provides *in vitro* and *in vivo* assays useful for determining the phenotypic outcome of a polymorphism in a candidate gene.

Every polymorphism has the potential to alter the genetic activity of an individual. At the level of a single gene, the effect of a polymorphism can range from an inconsequential, silent change to a change that causes a complete loss of protein function to a gain of aberrant or detrimental function mutation. The severity of the effect of a polymorphism on gene activity will depend on the exact

molecular consequences of the particular polymorphism. For example, alterations of a single pre-mRNA splicing dinucleotide could have profound effects on both the quantitative and qualitative properties of gene activity since alterations in splicing efficiency can both reduce the overall level of normal transcription as well as cause "exon skipping". If the deleted exon involves a coding exon then exon skipping will lead to an alteration in the amino acid composition of the resulting protein and likely effect protein activity. To accurately asses the role of a particular polymorphism in the regulation of various molecular events, appropriate assays for both gene expression and protein function must be carried out.

In vitro assays useful for determining the effects of a polymorphism on gene expression and protein function include, but are not limited to the following.

i. Transcriptional Regulation

The transcriptional regulation of a candidate gene containing a polymorphism may be altered, as compared to the wild type gene.

Promoter Activity

If a polymorphism is located in the promoter, enhancer or repressor region of a candidate gene, promoter assays (well known in the art) wherein the altered promoter of the candidate gene is used to drive the expression of a reporter gene (e.g. CAT, luciferase, GFP) are performed. Changes in the transcriptional regulation of a candidate gene due to the presence of a polymorphism can also be detected by methods useful for measuring the level of mRNA including S1 nuclease mapping and RT-PCR.

S1 Analysis

The S1 enzyme is a single-stranded endonuclease that will digest both single-stranded RNA and DNA. According to the method of S1 analysis, a probe that has been efficiently labeled to a high specific activity at the 5' end through the use of a kinase, is used to determine either the amount of an mRNA species or the 5' end of a message. A single stranded probe that is complementary to the sequence of the RNA species of interest is utilized in S1 analysis. If the structure of a particular mRNA species is known, S1 analysis is performed with oligonucleotide probes of at least 40 bp, that are complementary to the RNA of interest. It is preferable to use oligonucleotides wherein the 5' end of the oligonucleotide is complementary to the RNA. It is also preferable to use oligonucleotides wherein the 5' terminal residues contain dG or dC residues. If Si nuclease analysis will be utilized to determine the 5' termii of an RNA species, the 3' end of the oligonucleotide should extend at least 4

nucleotides beyond the RNA coding sequence. The inclusion of additional nucleotides facilitates differentiation of a band resulting from an RNA:DNA duplex and a band representing the probe.

A hybridization probe for S1 analysis is prepared by incubating 2pmol of an oligonucleotide in the presence of 150 mCi[y^{32}P]ATP (3000-7000Ci/mmol), 2.5 ml 10X T4 polynucleotide kinase buffer (700mM Tris-Cl, pH 7.5, 100 mM $MgCl_2$, 50 mM dithiothreitol, 1 mM spermidine-Cl, 1 mM EDTA), and l0U T4 polynucleotide kinase for 37^{0}C for 30-60 minutes. The radiolabeled probe is ethanol precipitated and resuspended at 1ml/0.3ng oligonucleotide or $1O^5$ cpm.

The hybridization reaction is performed as follows. An amount of probe equal to $5x10^4$ Cerenkov counts is added to 5Omg RNA on ice and ethanol precipitated. The resulting pellet is resuspended in 20ml S1 hybridization solution (80% deionized formamide, 40 mM PIPES, pH 6.4, 400mM NaCl, 1 mM EDTA, pH 8), denatured for 10 min at 65^{0}C and hybridized overnight at 30^{0}C. The following day, 300 ml of a mixture of 150 ml 2x S1 nuclease buffer (0.56M NaCl, 0.1 M sodium acetate, pH 4.5, 9mM $ZnSO_4$), 3ml 2mg/ml single-stranded calf thymus DNA, 147 ml H_20 and 300U S1 nuclease is added to the hybridization reaction and incubated for 60 minutes at 30^{0}C. Following the addition of 80ml S1 stop buffer (4M ammonium acetate, 20mM EDTA, 40 mg/ml tRNA) the sample is ethanol precipitated, resuspended in formamide loading dye, denatured and analyzed on a denaturing polyacrylamide/urea gel of the appropriate percentage for the expected size of the protected band (Ausubel et al., supra).

RT-PCR

The method of RT-PCR is useful according to the invention for RNA expression analysis. According to the method of reverse transcription /polymerase chain reaction (RT-PCR) during the reverse transcription (RT) step, the RNA is converted to first strand cDNA, which is relatively stable and is a suitable template for a PCR reaction. In the second step, the cDNA template of interest is amplified using PCR. This is accomplished by repeated rounds of annealing sequence- specific primers to either strand of the template and synthesizing new strands of complementary DNA from them using a thermostable DNA polymerase.

An RNA sample is ethanol precipitated with a cDNA primer. It may be preferable to use a cDNA primer that is identical to one of the amplification primers. To the pellet is added 12 ml H_20, 4ml 400mM TrisCl, pH 8.3, and 4 ml 400 mM KCl. The mixture is heated to 90^{0}C, slow cooled to 67^{0}C, microfuged and incubated for 3 hours at 52^{0}C. Following the addition of 29ml reverse transcriptase buffer (per sample/2.5ml 400mM TrisCl, pH8.3, 2.5ml 400mM KCl, 1ml 300mM $MgCl_2$, 5ml 100mM DTT, 5ml 5mM 4 dNTP mix, 2ml actinomycin D, 11ml H_20) and 0.5ml (16U) AMV

reverse transcriptase, the sample is incubated for 1 hour at a temperature between 37°C and 55°C. The temperature will be adjusted in accordance with the composition of the primer and the RNA of interest. The sample is then extracted sequentially with phenol and chloroform, and ethanol precipitated. The resulting cDNA pellet is resuspended in 40ml H_20. 5ml of the cDNA sample is mixed with 5ml or each amplification primer (~20mM each), 4ml 5mM 4dNTP mix, 10ml lOX amplification buffer (500mM KCl, 100mM TrisCi, pH8.4, 1mg/ml gelatin) and 70.5ml H_20. After the mixture is heated for 2 minutes at 94°C, 0.5 ml (2.5U) Taq DNA polymerase is added and the sample is overlaid with mineral oil. PCR amplification of the cDNA will be performed using the following automated amplification cycles: 39 cycles (2 minutes at 55°C, 2 minutes at 72°C, 1 minute at 94°C), 1 cycle (2 minutes at 55°C, 7 minutes at 72°C). The number of cycles can be varied in accordance with the abundance of RNA (Ausubel et al., supra).

If a polymorphism is located in a transcription factor binding site, assays including but not limited to the yeast two-hybrid assay (Fields et al., 1994, Trends Genet., 10:286) can be used to determine the effects of a polymorphism on transcription factor binding.

If the protein product of the gene of interest is a DNA binding protein the phenotypic outcome of a polymorphism may be impaired nuclear transport, DNA binding, chromatin assembly or chromatin structure, methylation or histone deacetylation.

Nuclear Transport

Immunocytochemical methods or cell fractionation techniques (as described above) are used to determine if the protein is correctly localized in the nucleus.

The DNA binding properties of a transcription factor are determined by gel shift analysis (as described in Ausubel et al., supra), oligonucleotide selection, southwestern assays or by immunohistochemical analysis of fixed chromosomes.

Gel Shift Analysis

The method of gel shift analysis is used to detect sequence specific DNA-binding proteins from crude extracts. According to this method, proteins that bind to an end-labeled DNA fragment will retard the mobility of the fragment. The change in the mobility of the labeled fragment is detected by the appearance of a discrete band comprising the DNA-protein complex.

A number of methods for preparing nuclear and cytoplasmic extracts useful for gel shift analysis are known in the art. For example, nuclear extracts are prepared according to the following method. A cell pellet is washed in PBS, resupended in a volume of hypotonic buffer (10 mM HEPES,

pH 7.9, 1.5 mM $MgCl_2$, 10mM KCl, 0.2 mM PMSF, 0.5 mM DTT) that is approximately equal to 3 times the packed cell volume and allowed to swell on ice for 10 minutes. Cells are homogenized in a glass Dounce homogenizer and the nuclei are collected by centrifugation and resupended in a volume of low-salt buffer (20 mM HEPES, pH 7.9, 25% (v/v) glycerol, 1.5 mM $MgCl_2$, 0.02 M KCl, 0.2 mM EDTA, 0.2 mM PMSF, 0.5 mM DTT) equivalent to one-half of the packed nuclear volume. Following the addition of a volume of high-salt buffer (20 mM HEPES, pH 7.9, 25% (v/v) glycerol, 1.5 mM $MgCl_2$, 1.2 M KCl, 0.2 mM EDTA, 0.2 mM PMSF, 0.5 mM DTT) equivalent to one-half of the packed nuclear volume (dropwise with stirring) to the nuclei, nuclear extraction is carried out for 30 minutes with continuous gentle stirring. The nuclei are collected by centrifugation and the nuclear extract is dialyzed against 50 volumes of dialysis buffer (20 mM HEPES, pH 7.9, 20% (v/v) glycerol, 100mM KCl, 0.2 mM EDTA, 0.2 mM PMSF, 0.5 mM DTT) until the conductivities of extract and buffer are equivalent. The extract is removed from the dialysis tubing and analyzed for protein concentration (Ausubel et al., supra).

Probes useful for gel shift analysis include a fragment of plasmid DNA or a gel-purified double stranded oligonucleotide. Preferably the probe is labeled with Klenow fragment by incubating a 1OOml solution of plasmid DNA or oligonucleotide with 1OOmCi of the desired [a-^{32}P] dNTP, 4ml of 5 mM 3dNTP mix and 2.5 U Klenow fragment for 20 minutes at room temperature. Upon the addition of 4ml of a solution comprising 5 mM of the dNTP corresponding to the radioactive dNTP, the sample is incubated for 5 minutes at room temperature. The radiolabeled probe is ethanol precipitated, resuspended in TE buffer and gel purified.

Gel shift analysis is performed by incubating 10,000 cpm of the labeled probe (0.1-0.5 ng) with 2mg poly (dI-dC)-poly(dI-dC), 300 mg BSA, and approximately 15mg of a nuclear extract or buffered crude protein extract prepared, for example, as described above, for 15 minutes at 30^{0}C. An aliquot of the binding reaction is analyzed by electrophoresis on a prewarmed low-ionic strength gel (e.g. a 4% polyacrylamide gel in TBE) and autoradiography (Ausubel et al., supra).

Oligoselection Assays for DNA Binding Activity

DNA binding activity is an essential property of proteins involved in many basic cell biological events, such as chromatin structure, transcriptional regulation, DNA replication and repair. The biological activity of a DNA binding protein can be assayed by defining the optimal target DNA binding site. Using the PCR based primer selection technique (Blackwell, 1990, Science, 250:1104) the canonical nucleotide sequence defining the binding site is elucidated *in vitro* by mixing purified full length protein, or just the DNA binding domain of a protein of interest, with an oligonucleotide duplex

pool containing a completely randomized central region flanked by primer-annealing sites. Multiple rounds of immunoprecipitation and amplification by PCR enriches for high affinity sites which are cloned are sequenced in order to define a canonical binding site.

The ability of a DNA binding protein to correctly regulate chromatin assembly and structure can be determined by DNase hypersensitivity assays. Alternatively, coimmunoprecipitation experiments or Western blot analysis can be used to determine if the DNA binding protein is associated with a component of the chromatin.

Southwestern Blot Assay for Protein-DNA Interactions

The ability of a protein to bind DNA is measured by using the "Southwestern" blot technique (for example see Antalis et al., 1993, Gene, 134:201). According to this method, radiolabelled DNA is incubated with protein that has been immobilized on nitrocellulose filters and the amount of bound DNA is measured by scintillation counting or autoradiography followed by densitometry. The protein to be tested can be pure protein, immunoprecipitated protein, crude cell lysates or even recombinant protein denatured directly from bacterial colonies, yeast or cell culture.

Assay of Protein Binding to Chromosomes *in Vivo*: Immunocytology of Fixed Chromosomes

Numerous biologically important nuclear proteins are in direct contact with genomic DNA. The presence of these proteins can be detected immunocytologically by fixing metaphase chromosomes such that the protein is permanently fixed at the region of DNA to which it normally binds. The presence and cytological location of the protein can then be determined by incubating the fixed chromosomes with an antibody directed against the protein of interest, and performing standard methods of immunohistochemical staining (Zink and Paro, 1989, Nature, 337:468).

Coimmunoprecipitation Assay for Chromatin Assemhly/Structure

If an antibody specific for a protein of interest exists, immunoprecipitation can be used to test for the presence of the protein (Otto and Lee, 1993, Methods Cell Biol., 37:119, Banting, 1995, In Gene Probes 1: A practical approach. Chapter 8: Antibody probes, pp. 225-227, IRL press.). The following methods are used for determining if a protein of interest is associated with a particular subcellular component. According to one method, proteins are immunoprecipitated with an antibody specific for a cellular component (e.g. chromatin or nuclear antigens), the immunoprecipitated material is analyzed on a gel by denaturing polyacrylamide gel electrophoresis and western blot analysis is performed with an antibody specific for the protein of interest, to determine if a physical association

exists between the cellular component and the protein of interest. Various incubation and wash treatments of the cell lysate are used to remove background contamination and enhance the sensitivity of detection (Banting, 1995, supra). Alternatively, the initial immunoprecipitation can be carried out with the antibody specific for the protein of interest, and the western blot analysis can be performed with an antibody specific for a cellular component. According to a variation of this method, prior to immunoprecipitation the cells can be treated with a protein crosslinker to ensure that protein-protein interactions are maintained during immunoprecipitation. According to another variation of this method, proteins can be cross-linked to DNA and then precipitated (Dedon et al., 1991, Anal. Biochem., 197:83). If DNA coprecipitates with a particular protein, this suggests that DNA is associated with, and presumably bound to the protein. The coprecipitating DNA can be sequenced to identify the bound sequence.

DNAse Hypersensitivity

The transcriptionally active promoter region of a gene can be analyzed for susceptibility to cleavage by DNAseI (Montecino et al., 1994,Biochemistry, 33:348). Efficient cleavage of genomic DNA is dependent on the accessibility of this enzyme to the DNA, and is influenced by several factors, including nucleosome packaging, overall chromatin configuration, and the presence of DNA binding proteins such as transcription factors. DNA sequence variations within the promoter DNA may have profound effects on these factors and result in aberrant regulation of gene transcription and ultimately abnormal biological activity of the gene. Therefore, altered gene activity around a polymorphic site can be detected as increased or decreased DNAseI hypersensitivity (Vaishnaw et al., 1995, Immunogenetics, 41:354).

Assay for DNA Methylation

Accurate mapping of DNA methylation patterns, for example, in CpG islands which are unmethylated regions of DNA, is used to investigate and gain a better understanding of diverse biological processes such as the regulation of imprinted genes, X chromosome inactivation and tumor suppressor gene silencing in human cancer. DNA methylation at specific sites is most frequently studied by use of methylation-sensitive restriction endonucleases (for example HpaII) and Southern blotting (Sambrook et al., supra). The sensitivity of this method can be enhanced several hundred-fold by performing a ligation-mediated PCR step (as described in Steigerwald et al., 1990, Nucleic Acids Res., 6:1435) after enzyme treatment. An alternative strategy termed methylation-specific PCR (Herman et al., 1996, Proc Natl Acad Sci USA., 93:9821), is used to determine the methylation status

of CpG islands without the use of methylation-specific restriction enzymes.

Histone-Deacetylation

Transcription of chromatin-packaged genes involves highly regulated changes in nucleosome structure that control DNA accessibility. Changes in nucleosome structure can be mediated by enzymatic complexes which control the acetylation and deacetylation of histones. Transcription elongation is required for the formation of the unfolded structure of transcribing nucleosomes, and histone acetylation is required for the maintenance of these structures (Walia et al., 1998, J. Biol. Chem., 3:14516). Deacetylation can be prevented by incubating cells with histone deacetylase inhibitors such as sodium butyrate or trichostain A. To assay for changes in acetylation and the state of transcriptional activity, chromatin fractions are purified using organomercury and hydroxylapatite dissociation chromatographic techniques (Walia et al., supra).

ii. Transcription Start Site

To determine if a particular polymorphism causes a change in the transcriptional start site of a candidate gene S1 nuclease mapping and primer extension can be performed. The presence of a polymorphism may cause an mRNA to be aberrantly expressed. In particular, a polymorphism may change the tissue specificity or developmental expression pattern of an mRNA species. A variety of molecular methods for detecting mRNA known in the art can be performed to determine the expression pattern of an mRNA These methods include, but are not limited to the following: Northern blot analysis, RT-PCR, S1 analysis, RNase Protection analysis, or *in situ* hybridization analysis of sections, wherein the samples are derived from multiple different tissues or from a tissue at different stages of development. Northern blot analysis, RT-PCR and S1 analysis can also be used to determine if a polymorphism results in an altered pattern of mRNA splicing.

Northern-Blotting

The method of Northern blotting is well known in the art. This technique involves the transfer of RNA from an electrophoresis gel to a membrane support to allow the detection of specific sequences in RNA preparations.

Northern blot analysis is performed according to the following method. An RNA sample (prepared by the addition of MOPS buffer, formaldehyde and formamide) is separated on an agarose/formaldehyde gel in 1X MOPS buffer. Following staining with ethidium bromide and

visualization under ultra violet light to determine the integrity of the RNA, the RNA is hydrolyzed by treatment with 0.05M NaOH/1.5MNaCl followed by incubation with 0.5M Tris-Cl (pH 7.4)/1.5M NaCl. The RNA is transferred to a commercially available nylon or nitrocellulose membrane (e.g. Hybond-N membrane, Amersham, Arlington Heights, IL) by methods well known in the art (Ausubel et al., supra, Sambrook et al., supra). Following transfer and UV cross linking, the membrane is hybridized with a radiolabeled probe in hybridization solution (e.g. in 50% formamide/2.5% Denhardt's/100-200mg denatured salmon sperm DNA/0. 1% SDS/5X SSPE) at 42^{0}C. The hybridization conditions can be varied as necessary as described in Ausubel et al., supra and Sambrook et al., supra. Following hybridization, the membrane is washed at room temperature in 2X SSC/0.1% SDS, at 42^{0}C in 1X SSC/0.1% SDS, at 65^{0}C in 0.2X SSC/0.1% SDS, and exposed to film. The stringency of the wash buffers can also be varied depending on the amount of background signal (Ausubel et al., supra).

RNase Protection Analysis

RNase Protection analysis can be used to analyze RNA structure and amount and determine the endpoint of a specific RNA.

The method of RNase protection is more sensitive than S1 analysis since it utilizes a sequence specific hybridization probe that is labeled to a high specific activity. The probe is hybridized to sample RNAs and treated with ribonuclease to remove free probe. Following ribonuclease treatment, the fragments comprising probe annealed to homologous sequences in the sample RNA are recovered by ethanol precipitation, and analyzed by electrophoresis on a sequencing gel. The presence of the target mRNA is indicated by the presence of an appropriately sized fragment of the probe.

A probe is labeled by the method of *in vitro* transcription (in the presence of [a-^{32}P] CTP as described in Section B entitled "Production of a Polynucleotide Sequence". The RNA sample to be analyzed is ethanol precipitated and resuspended in 30ml hybridization buffer (4 parts formamide/1 part 200 mM PIPES, pH 6.4, 2 M NaCl, 5 mM EDTA) containing 5 x 10^5 cpm of the probe RNA. The mixture is denatured 5 minutes at 85^{0}C and incubated at the desired hybridization temperature (30^{0}C to 60^{0}C) for >8 hours. To each reaction mixture is added 350 ml ribonuclease digestion buffer (10 mM Tris-Cl, pH 7.5, 300 mM NaCl, 5 mM EDTA) containing 40 mg/ml ribonuclease A and 2 mg/m1 ribonuclease T1. The sample is incubated for 30-60 minutes at 30^{0}C. Following the addition of 10 ml 20%SDS and 2.5ml 20 mg/ml proteinase K, the sample is incubated for 15 minutes at 37^{0}C. The sample is extracted with phenol /chloroformlisoamyl alcohol, ethanol precipitated, resuspended in RNA loading buffer (80% (v/v) formamide, 1 mM EDTA, pH 8.0, 0.1 % bromophenol blue, 0.1 % xylene

cyanol), denatured and analyzed by electrophoresis on a denaturing polyacrylamide/urea gel and autoradiography (Ausubel et al., supra).

Primer Extension

The method of primer extension is used to map the 5' end of an RNA and to quantitate the amount of an RNA of interest by using reverse transcriptase to extend a primer that is complementary to a region of a given RNA.

An oligonucleotide primer is labeled in a kinase reaction as described for S1 analysis. The primer extension reaction is performed by mixing 10-50 mg total cellular RNA (in 10ml) with 1.5ml 10X Hybridization buffer (1.5M KCl, 0.1M TrisCl, pH 8.3, 10mM EDTA) and 3.5 ml labeled oligonucleotide. Samples are heated to 65^{0}C for 90 minutes and allowed to slow cool at room temperature. To each sample is added 30 ml of primer extension reaction mixture (0.9 ml Tris-Cl, pH 8.3, 0.9 ml 0.5M $MgCl_2$, 0.25 ml DTT, 6.75 ml 1 mg/ml actinomycin D, 1.33 ml 5 mM 4dNTP mix, 20 ml H_20, 0.2ml 25 U/ml AMV reverse transcriptase). Samples are incubated for 1 hour at 42^{0}C, and then, following the addition of 105 ml RNase reaction mix (100 mg/ml salmon sperm DNA, 20 mg/ml RNase A) for 15 minutes at 37^{0}C. Samples are extracted in phenol/chloroformlisoamyl alcohol, ethanol precipitated, resuspended in stop/loading dye (20 mM EDTA, pH 8.0, 0.05% bromophenol blue, 0.05% xylene cyanol in formamide), heated at 65^{0}C and analyzed by electrophoresis on a 9% acrylamide/7M urea gel and autoradiography.

In Situ Hybridization

Cytological techniques well known in the art can be used to determine the temporal and spatial expression patterns of mRNA (*in situ* hybridization of tissue sections) and protein (immunohistochemistry in individual cells).

Preparation of histological samples

Tissue samples intended for use in *in situ* detection of either RNA or protein are fixed using conventional reagents; such samples may comprise whole or squashed cells, or sectioned tissue. Fixatives useful for such procedures include, but are not limited to, formalin, 4% paraformaldehyde in an isotonic buffer, formaldehyde (each of which confers a measure of RNAase resistance to the nucleic acid molecules of the sample) or a multi-component fixative, such as FAAG (85 % ethanol, 4% formaldehyde, 5% acetic acid, 1% EM grade glutaraldehyde). For the detection of RNA, water used in the preparation of an aqueous component of a solution to which the tissue is exposed until it is

embedded is RNAase-free, i.e. treated with 0.1% diethylprocarbonate (DEPC) at room temperature overnight and subsequently autoclaved for 1.5 to 2 hours. Tissue will be fixed at 4^{0}C, either on a sample roller or a rocking platform, for 12 to 48 hours in order to allow the fixative to reach the center of the sample.

Prior to embedding, excess fixative will be removed and the sample will be dehydrated by a series of two- to ten-minute washes in increasingly high concentrations of ethanol, beginning at 60% and ending with two washes in 95% and another two in 100% ethanol, followed by two ten-minute washes in xylene. Samples will be embedded in one of a variety of sectioning supports, e.g. paraffin, plastic polymers or a mixed paraffin/polymer medium (e.g. Paraplast®Plus Tissue Embedding Medium, supplied by Oxford Labware). For example, fixed, dehydrated tissue will be transferred from the second xylene wash to paraffin or a paraffin/polymer resin in the liquid-phase at about 58^{0}C. The paraffin or a paraffin/polymer resin will be replaced three to six times over a period of approximately three hours to dilute out residual xylene. The sample will be incubated overnight at 58^{0}C under a vacuum, in order to optimize infiltration of the embedding medium into the tissue. The next day, following several additional changes of medium at 20 minute to one hour intervals, also at 58^{0}C, the tissue sample will be positioned in a sectioning mold, the mold will be surrounded by ice water and the medium will be allowed to harden. Sections of 6mm thickness will be taken and affixed to 'subbed' slides, which are slides coated with a proteinaceous substrate material, usually bovine serum albumin (BSA), to promote adhesion. Other methods of fixation and embedding are also applicable for use according to the methods of the invention; examples of these are found in Humason, G.L., 1979, Animal Tissue Techniques, 4th ed. (W.H. Freeman & Co., San Fransisco), as is frozen sectioning (Serrano et al., 1989, supra).

In situ Hybridization Analysis

According to the method of *in situ* hybridization a specifically labeled nucleic acid probe is hybridized to cellular RNA present in individual cells or tissue sections. *In situ* hybridization can be performed on either paraffin or frozen sections. Depending on the desired sensitivity and resolution, either film or emulsion autoradioagraphy can be utilized to detect the hybridized radioactive probe.

The following method of *in situ* hybridization is performed by incubating slides containing cell or tissue specimens in a slide rack contained within a glass staining dish. According to this method, it is preferable to use solutions that have been prepared fresh. Prior to the hybridization steps, slides are dewaxed to remove the sectioning support material. The dewaxing protocol involves sequential washes in xylene, rehydration by sequential washes in 100%, 95%, 70% and 50% ethanol, and

denaturation in 0.2N HCl. Following a heat denaturation step (70^{0}C in 2X SSC), samples are postfixed in a freshly prepared solution of 4% PFA, washed in PBS, incubated in 10 mM DTT (10 min at 45^{0}C) and blocked in 400 ml PBS containing 0.617g DTT, 0.74 g iodoacetamide and 0.Sg N-ethylmaleimide, for 30 min at 45^{0}C in a water bath covered with aluminum foil, due to the light sensitivity of iodoacetamide and N-ethylmaleimide. The samples are washed in PBS and equilibrated sequentially in freshly prepared 0. 1M triethanolamine (TEA buffer), TEA buffer/0.25% acetic anhydride, and TEA buffer/0.5% acetic anhydride. Following a blocking step in 2X SSC, the sample are dehydrated by sequential washes in 50%, 70%, 95%, and 100% ethanol and air dried. ^{35}S-labeled riboprobes and competitor probes prepared in the absence of a radiolabel (prepared as described in Section B entitled "Production of a Polynucleotide Sequence") or double-stranded DNA probes (prepared with [^{35}S]dNTPs by methods well known in the art including nick translation or random oligonucleotide-primed synthesis) are heated to 100^{0}C for 3 min and diluted to a concentration of 0.3 mg/ml final probe concentration, in 50% formamide, 0.3M NaCl, 10mM TrisCl, pH 8.0, 1 mM EDTA, lx Denhardt solution, 500 mg/ml yeast tRNA, 500 mg/ml poly(A) (Pharmacia), 50 mM DTT, 10% polyethylene glycol (MW 6000). The hybridization step is carried out by covering the sample with an appropriate amount of probe, and incubating for 30 min to 4 hour at 45^{0}C in a chamber designed to prevent dilution or concentration of the hybridization solution. Samples are washed sequentially at 55^{0}C in solution A (50% (v/v) formamide, 2X SSC, 20 mM 2-mercaptoethanol), and solution B (50% (v/v) formamide, 2X SSC, 20 mM 2-mercaptoethanol, 0.5% (v/v) Triton-X-100) and at room temperature in solution C (2X SSC, 20 mM 2- mercaptoethanol). Following a 15 minute incubation with RNase, samples are washed at 50"C in solution C, and at room temperature in 2X SSC. Samples are rehydrated by sequential washes in 50% ethanol/0.3M ammonium acetate, 70% ethanol/0.3M ammonium acetate, 95% ethanol/0.3M ammonium acetate, and 100% ethanol. Slides are air dried and analyzed by film or by emulsion autoradiography (Ausubel et al., supra).

iii. mRNA Stability/Control of Turnover and mRNA Transcription Rate

Changes in mRNA stability/control of turnover and mRNA transcription rates due to the presence of a polymorphism, can be detected by the following methods.

<u>mRNA Stability</u>

Gene-expression can be regulated by variations in mRNA stability (Liebhaber, 1997, Nucleic Acids Symp Ser., 36:29 and Ross J. 1996, Trends Genet., 5:171). Any gene variation occurring within

the cis-acting elements which control mRNA abundance may influence gene expression levels (Peltz et al., 1992, Curr Opin Cell Biol., 4:979). Quantitative RT-PCR (Kohler, et al, 1995, Quantitation of mRNA by polymerase chain reaction, Springer) and mRNA radiolabelling techniques are two methods for measuring relative mRNA abundance and stability. Quantitative PCR employs an internal standard to provide a direct comparison between alternative reactions, enabling comparison of low abundance transcripts or transcripts derived from a sample that is only available in a limited quantity (McPherson MJ et al., eds, 1995, PCR2- A practical approach. IRL Press).

Assay for mRNA Transcription Rates

Genetic polymorphism within the regulatory regions of a gene can significantly alter transcription rate and mRNA stability, resulting in reduced biological activity of the encoded protein. One of the most sensitive assays for measuring the rate of gene transcription is the nuclear runoff assay (Groudine and Casimir, 1984, Nucleic Acids Res 12: 1427). Nuclei isolated from cell lines expressing the target gene of interest are treated with radiolabelled UTP and the level of incorporation of radiolabel into nascent RNA transcripts is determined by filter hybridization to immobilized cDNA derived from the target gene.

iv. Intracellular mRNA Localization

A genetic variation can cause a change in the localization of a particular mRNA species (e.g. to the cytoskeleton, or to the nuclear scaffold).

Immunohistochemisitry

Changes in RNA localization can be detected by immunohistochemical methods well known in the art (e.g. *in situ* analysis described above).

Oocyte Injection Assays

In many cases mRNA, like protein, is localized in relation to the polarity of the cell or the cytoskeletal architecture (St. Johnston, 1995, Cell, 81:161). The *Xenopus* oocyte is a popular, experimentally tractable, system for studying intracellular trafficking of mRNA (Nakielny et al., 1997, Annu. Rev. Neurosci., 20:269). Fluorescently labelled RNA is microinjected into the large oocyte cell where its location can be detected using standard microscopy methods. Polymorphic variants of a particular mRNA species may differ in their response to cellular mechanisms responsible for partitioning mRNA within the cell. This method has been useful for demonstrating that sequence

variations can affect sub-cellular localization (Grimm et al., 1997,EMBO J., 16:793)

v. Post-Translational Alterations

Post-Translational alterations resulting from premature stop codons, translational readthrough or multiple open reading frames and translational suppression may occur as a result of a polymorphism. To detect post-translational alterations, a polynucleotide comprising one or more polymorphisms is subjected to *in vitro* transcription and *in vitro* translation (as described in sections B and J entitled "Production of a Polynucleotide Sequence" and "Preparation of a Labeled Protein"). The translation product(s) are analyzed for the appearance of aberrantly sized proteins. Additional post-translational alterations that may occur as a result of a polymorphism include changes in localization due to an altered signal sequence, and changes in glycosylation, myristilation, and susceptibility to or sites of proteolytic cleavage.

The method of immunocytochemistry can be used to determine if a protein is incorrectly localized, due to the presence of an altered signal sequence.

Immunohistochemistry

Immunohistochemical techniques including indirect immunofluorescence, immunoperoxidase labeling or immunogold labeling, are used for protein localization.

Immunofluorescent labeling of tissue sections (prepared as for *in situ* analysis, described above) is performed by the following method. Slides containing the sample of interest are equilibrated to room temperature washed in PBS, incubated with an appropriate dilution of primary antibody (1 hour at room temperature), washed in PBS, incubated with an appropriate dilution of secondary antibody (1 hour at room temperature), washed in PBS and analyzed under a microscope (Ausubel et al., supra). Alternatively, the sensitivity of the immunohistochemical reaction is increased by using a streptavidin-secondary antibody conjugate reacted with a biotinfluorochrome conjugate. Alternatively, immunogold labeling is used to detect a protein of interest by using an immunogold-conjugated secondary antibody.

Immunoperoxidase labeling of tissue sections is performed by the following method. Slides are pretreated in 0.25% hydrogen peroxide, incubated with primary antibody, washed in PBS and incubated (1 hour at room temperature) with a specific secondary bridging antibody capable of recognizing both the primary antibody and a Horseradish peroxidase antiperoixidase (PAP) complex. The slides are washed in PBS and developed in diaminobenzidene substrate solution (0.03% (w/v) 3,3' diaminobenzidene in 200 ml PBS) at room temperature (Ausubel et al., supra).

Alternatively, protein localization is determined by cell fractionation wherein cells are biosynthetically labeled, the labeled material is fractionated, and the radiolabeled proteins in each fraction are analyzed by immunoprecitation with an antibody specific for the protein of interest.

Assay for Glycosylation Inhibition

Changes in protein glycosylation can be detected by radiolabelling a protein of interest with sugars, determining if a change in the cellular localization (by immunocytochemistry) of the protein in culture has occurred due to aberrant glycosylation, or by determining the effects of inhibitors of glycosylation on the migration pattern of proteins analyzed by polyacrylamide gel electrophoresis.

Post-translational glycosylation of proteins plays an important role in defining protein function (Baeziger, 1994, FASEB J., 13:1019; Jacob, 1995, Curr. Opin. Struct. Biol., 5:605). Protein glycosylation can be inhibited by tunicamycin, an antibiotic, as well as by several sugar analogues (Schwarz, 1991, Behring Inst Mitt., 89:198). These reagents are used to characterize the effects of sequence changes on protein glycosylation.

Assay for Post-Translational Modification with Lipids

Changes in protein modification with lipids (e.g. myristilation) are detected by radiolabelling a protein of interest with myristic acid or by determining if a change in the cellular localization of the protein in culture has occurred as a result of aberrant lipid modification (by immunocytochemistry).

Covalent attachment of lipids is a mechanism by which eukaryotic cells direct and, in some cases, control, membrane localization of proteins (Casey, 1994, Curr. Opin. Cell. Biol., 2:219). Such post-translational addition of myristyl, palmityl or prenyl side-chains has a key role in the functional regulation of many proteins (Chow et al., 1992, Curr. Opin. Cell. Biol., 4:629; Resh, 1994, Cell, 763:411). Assays for detecting proteins that are covalently modified by the attachment of lipids include labeling with [^{3}H]myristate (Stevenson et al., 1992, J. Exp. Med., 176:1053), or a combination of enzymatic and chemical cleavage techniques performed in conjunction with tandem mass spectrometry to determine sites of modification (Papac et al., 1992, J. Biol. Chem., 267:16889).

Proteolytic Cleavage

Post-translational cleavage of polypeptides is an important mechanism for modulating protein function in many physiological processes. Protease activity is involved in zymogen processing, activation of enzyme catalysis, tissue/cell remodeling, signal transduction cascades, protein degradation and cell death pathways (Rappay, 1989, Prog Histochem Cytochem., 18:1). A protein that is predicted

to be a protease or the target of a protease can be assayed *in vitro* using purified proteins or cell extracts (Muta et al., 1995, J. Biol. Chem. 270:892) where cleavage efficiency is monitored by standard PAGE or western blotting. Alternatively, proteases and/or their targets can be expressed from expression plasmids in *in vivo* cell culture systems in order to monitor their biological activity (Zhang, et al., 1998, J. Biol. Chem. 273:1144). The specificity of proteolytic cleavage is determined using inhibitors that selectively block seine, cysteine, aspartic and metallo proteolytic activity (e.g. pepstatin A selectively inhibits aspartic proteases) (Rich, et al., 1985, Biochemistry., 24: 3165).

To determine if a protein has been modified such that the sites of proteolytic cleavage have been altered, or susceptibility to proteolytic cleavage has changed pulse chase experiments with radiolabeled protein can be carried out to determine the precursor-product relationship following digestion with a protease of a given specificity. The method of pulse chase labeling is described in Ausubel et al., supra. Alternatively, inhibitors of proteases (e.g acid proteases or seine proteases) can be used to identify protease cleavage sites.

vi. Changes in Receptor Properties

If the gene of interest encodes a receptor protein, a polymorphism may modify the properties of the receptor such that receptor binding/turnover or activation is altered. Receptor formation can be impaired if a polymorphism causes improper receptor localization or assembly.

Receptor Localization

To determine if a receptor protein is being expressed at the proper location (e.g. nucleus, cytoplasm, cell surface), the receptor can be localized by immunocytochemical techniques. Alternatively, cells that are expressing the receptor can be fractionated and subjected to Western blot analysis or biosynthetically labeled, fractionated and analyzed by immunoprecipitation.

Protein-Protein Interactions/*In vitro* Assembly Assays for Receptors

A number of methods can be used to determine if a receptor is colocalized with the appropriate protein partner.

The function of a protein may be dependent on the ability of the protein to interact with other proteins as part of a large complex. For example, certain cell surface receptors consist of a receptor complex that is composed of several homo- or heteromeric protein subunits, and activation by ligand can result in altered protein-protein interactions both within the receptor complex and with "downstream" targets such as G-proteins (Okada and Pessin, 1996, J. Biol. Chem., 271:25533).

Protein-protein interactions can be assayed immunologically by coimmunoprecipitation of native (Gilboa et al., 1998, J. Biol. Chem., 140:767) or chemically cross-linked complexes (Haniu et al., 1997, J. Biol. Chem., 272:25296), or through protein-protein mobility shift assays (Stern and Frieden, 1993, Anal. Biochem., 212:221). If all of the components of a receptor complex have been identified, one can employ *in vitro* reconstitution assays to assess whether a single protein alteration can effect the functioning of the entire complex (Durovic et al., 1994, J. Biol. Chem., 269:30320).

Assay for *In Vitro* Assembly of Multimeric Protein Complexes

To determine whether these genetic variations have affected protein complex assembly, experiments are carried out wherein recombinant mutant subunits are transfected into cells and coexpressed with the other subunit components *in vitro*. Proper assembly is assessed by immunoprecipitation of the protein complex in question with antibodies specific for the various members of the complex followed by PAGE analysis (Koster et al., 1998, Biophysl. J., 74:1821).

Assay Receptor Binding/Turnover

Receptor-ligand interaction is essential for the functionality of the bound complex. Genetic changes that alter either ligand or receptor can dramatically affect receptor binding, turnover, and subsequent activation of downstream signaling events. Receptor binding/turnover can be measured by standard Scatchard analysis of radiolabelled ligand binding *in vitro* (Culouscou et al., 1993, J. Biol. Chem. 268:10458) or in cellular based assays (Greenlund et al., 1993, J. Biol. Chem. 268: 18103).

Ligand Binding as Measured by Affinity Chromatography

Alternatively, affinity chromatography methods (well known in the art) can be employed to determine if a receptor is demonstrating aberrant binding characteristics. According to the method of affinity chromatography, receptor-ligand interactions are allowed to occur, and the binding efficiency or receptor and ligand and/or turnover of receptor-ligand complexes is measured. Alternatively, affinity chromatography can be used to isolate one or more components of a receptor ligand interaction for further analysis (March et al., 1974, Adv. Exp. Med. Biol., 42:3). The method of affinity chromatography typically involves immobilizing on a solid support one component, for example a known ligand for a receptor, and then incubating the immobilized ligand with radiolabelled protein under optimal binding conditions. To measure the exact binding affinity of a given ligand-receptor pair, an increasing amount of non-labeled competitor is added. This assay can be used to assess altered binding efficiency resulting from the presence of a polymorphism in a protein of interest.

Receptor Activation Assays: Phosphorylation, Kinase Activity and Mitogenic Stimulation

Almost all signaling that occurs through cell surface receptors is regulated by phosphorylation, a reversible post-translational event that occurs at specific amino acid residues and is catalyzed by a protein kinase activity present within the receptor itself (autophosphorylation) or in trans via direct interaction with an associated kinase (Hunter, 1997, Philos Trans R Soc Lond B Biol Sci., 353:583). The specific effect of phosphorylation on a biological activity depends on the receptor, but often results in modulation of endogenous receptor kinase activity or interaction with associated proteins, which are also often kinases. The results of a phosphorylation event are passed on through a cascade of protein kinases/phosphatases which ultimately effect downstream processes controlling gene transcription, cell proliferation, metabolism, movement and differentiation (Patarca, 1996, Crit Rev Oncog., 7:343). The biological function of a receptor is usually assayed in cell culture following over-expression. The phosphorylated state of a receptor can be assayed directly by immunological methods by employing an antibody that specifically recognizes a phosphorylated residue (Bangalore, 1992., Proc Natl Acad Sci USA., 89:11637). Endogenous kinase activity associated with a receptor is measured via the incorporation of radiolabelled phosphate in immunoprecipitated receptor complex (Kazlauskas and Cooper, 1989, Cell 58:1121). "Downstream" events of receptor activity including mitogenic stimulation or map kinase activity, can be measured by tritiated thymidine incorporation (Luo et al., 1996, Cancer Res. 56:4983), or by mobility-shift analysis of map kinase on western blots (Vietor, 1993., J. Biol. Chem. 268:18994), respectively.

Immunocytochemical methods can be used to determine if a receptor-ligand complex is correctly translocated to the nucleus. Alternatively, nuclear preparations (prepared as described below) can be analyzed by Western blot or immunoprecipitation for the presence of the receptor protein.

If a receptor is a transcriptional activator, the ability of the receptor to induce gene expression can be measured by a variety of methods including Northern blot analysis, or reporter gene assays wherein the promoter region isolated from a gene that is activated by the receptor regulates the expression of a reporter protein.

vii. Enzyme Catalysis

The gene of interest may encode a protein that has an enzymatic activity wherein the enzyme catalyzes a reaction that is critical to the general metabolism of a cell. To determine if a mutated protein is impaired in its enzymatic function, assays can be performed to measure the enzymatic activity of the protein. There are many important enzymatic activities associated with normal cellular

metabolism, including: glycosidation, esterification, amidation, hydroxylation, acetylation, sulfonylation, alkylation. Each of these activities are assayed using *in vitro* methods employing overexpressed or purified proteins, well known in the art (Eisenthal and Danson, 1992, Enzyme Assays: A Practical Approach, Rickwood et al., Eds., IRL Press. Oxford, England).

The protein of interest may also be involved in various aspects of DNA synthesis or replication. *In vitro* assays for the enzymatic reactions involved in DNA synthesis or replication (e.g. polymerase, ligase, exonuclease or helicase activity) are known in the art. The biological activity of the proteins catalyzing these activities are assayed *in vitro* using standard enzymatic techniques (Adams, 199, DNA Replication: A Practical Approach I, Rickwood, et al., Eds., IRL Press. Oxford, England).

If the protein of interest is involved in glycolysis or energy transport, assays for measuring transporter activity or the activity of ATP dependent pumps are useful, according to the invention, for determining if a mutated protein is impaired in these functions.

Transporter Activity

Mammalian cells possess a variety of transporter systems, for example amino acid transporters, which have overlapping substrate specificity (Van Winkle et al., 1993, Biochim Biophys Acta, 1154:157). To determine if a polymorphism in a candidate gene of interest has altered the function of the protein product of this gene as a molecular transporter, the full-length cDNA clone is isolated by standard expression cloning strategies, and a change in activity of the full-length cRNA or antisense cRNA upon microinjection into Xenopus laevis oocytes is determined by measuring changes in influx/efflux transport of radiolabelled amino acid molecules (Broer et al., 1995, Biochem J., 312(Pt 3):863), neurotransmitters or their metabolites.

ATP-dependent pumps Activity

Mammalian cells possess a variety of molecules that are categorized as ATP-binding cassette or ATP-dependent transporters or pumps. These include the Na^+-K^+-ATPase ion pump, the calcium uptake pump, (K^+ + H^+)-ATPase and the human multidrug resistant protein termed P-glycoprotein. Alterations in pump activity are investigated by expressing the clone specific for the pump protein(s) of interest in *Xenopus* oocytes, and performing tracer studies which measure the changes in ATP-dependent uptake or extrusion of a radiolabelled substrate, and changes in the coupling ratios (e.g. moles substrate transported/mole ATP hydrolyzed) (Shapiro et al., 1998, Eur. J. Biochem., 254:189).

viii. Ion Channel

The gene of interest may encode for a protein that is a component of an ion channel. Immunocytochemical methods can be used to determine if an ion channel protein demonstrates the appropriate cell type specificity.

The activity of an ion channel can be measured by electrophysiological methods in oocytes. Alternatively, the sensitivity of ion channel activity to a particular inhibitor can be determined.

Assays for Ion Channel Activity in Oocytes

Polymorphisms which alter ion channel function and regulation are studied using the oocytes of *Xenopus laevis*. Injection of the oocytes with exogenous *in vitro* transcribed mRNA results in the production and functional expression of foreign membrane proteins, including voltage- and neurotransmitter- operated ion channels (Dascal et al., 1987., CRC Crit Rev Biochem., 224:317). Changes in the oocyte transmembrane current in response to expression of an exogenous mRNA is measured. This technique has been improved by the development of rapid superfusion systems that utilize a dual role perfusion micropipette that controls internal solution as well as monitoring voltage (Costa et al., 1994, Biophys J., 67:395). This technology represents a useful system for studying various aspects of ion channels encoded for by foreign mRNAs including channel expression, single-channel behavior, and the response of channels to the action of pharmacologically active substances (Sigel, 1987,J. Physiol., 386: 73).

Patch Clamp Assays for Ion Channel Activity

The function of individual channel proteins is determined by the high resolution patch clamp technique. This technique (which is useful in a variety of cell types, including *Xenopus* oocytes described above) involves measuring changes in transmembrane current across the cell membrane *in vitro* (Sachs et al., 1983, Methods Enzymo1., 103: 147). Processes such as signaling, secretion, and synaptic transmission are examined at the cellular level by the patch clamp method. The gene expression pattern and protein structure of ionic channels can be determined by combining information derived from high-resolution electrophysiological recordings obtained by the patch clamp method with molecular biological analysis (Liem et al., 1995, Neurosurgery, 36: 382).

A polymorphic variation in a gene that encodes a protein that is a member of a multimeric protein complex, such as an ion channel or a cytoskeletal structural component, can alter the assembly and function the multimeric protein complex (Lee et al., 1994., Biophys J., 66: 667). A gene variation may affect protein-protein interaction, or disrupt the production of components of a multimeric complex, thereby disrupting stoichiometry and consequently decreasing stability.

Assay for *In Vitro* Assembly of Multimeric Protein Complexes

In vitro assembly assays (described above) can be performed to determine if a polymorphism has affected the assembly of an ion channel.

ix. Cellular Properties

The influence of a polymorphism on general aspects of cell behavior, including cell morphology, adhesive properties, differentiation and proliferation can be assessed using a combination of methods including microscopic observation of cell cultures (Azuma et al., 1994, Histol.Histopathol., 9:781), immunohistochemistry, and FACs analysis techniques (Beesley, 1993, Immunocytochemistry: a Practical Approach, Rickwood, et al., (Eds), IRL Press and Ormerod, 1994, Flow Cytometry: a practical Approach, Rickwood et al., (Eds), IRL Press. Oxford, England).

Assays for Measuring Apoptosis

Apoptosis has been implicated in the etiology and pathophysiology of a variety of human diseases. Gene variants which influence the process of apoptosis can be assessed by a variety of methods of analysis involving either the tissues or cells (Allen et al., 1997, J Pharmacol Toxicol Methods, 37: 215). Cell cultures expressing the gene variants of interest are analyzed using Annexin V which interacts strongly with phosphatidylserine residues that have been exposed as a result of plasma membrane breakdown occurring in the early stages of apoptosis. Either vital or fixed material can be analyzed by Annexin V labeling in combination with microscopy and flow cytometry detection methods (van Engeland et al., 1998, Cytometry, 31:1). TdT-mediated deoxyuridine triphosphate (dUTP)-biotin nick end-labeling (TUNEL) is a preferred method for specific staining of apoptotic cells in histological sections and cytology specimen (Labat-Moleur et al., 1998, J. Histochem Cytochem., 46:327; Sasano et al., 1998., Diagn Cytopathol.,18:398). Apoptosis is also detected by quantification of DNA fragmentation by ethidium bromide staining and gel electrophoresis, or by the use of saturation labeling of 3' ends of DNA fragments (Peng and Liu, 1997, Lab Invest., 77:547).

Assay for *In Vivo* Receptor Function: Growth Cone Guidance Assay

Activation of cell-surface receptors can result in the stimulation of cell motility. There are many different families of signaling molecules, for example the netrins, (Serafini et al., 1994, Cell. 78: 409), which are responsible for both contact mediated or chemo-mediated attraction and repulsion of migrating cells. A classic model for this activity is the trajectory that the leading edge "growth cone" takes when a neuron is stimulated to grow out from explanted neural tissue in cell culture (Goodman,

1996, Annu Rev Neurosci. 19: 341). Ligands present in the culture medium or immobilized on a substrate bind to receptors on the cell-surface of the growth cone and trigger second-messenger signals thereby dictating an appropriate steering response. The biological activity of such receptors or ligands can be measured by overexpressing the receptor or ligand protein in culture and then monitoring growth cone guidance (Kremoser et al., 1995, Cell 82: 359). Attraction or repulsion of cells which is observed to be different than normal is an indication of the role of this protein in growth guidance, and identifies the polymorphisms as altering function.

x. Changes in gene expression or protein function that result from the presence of a polymorphism can be detected by *in vivo* assays including the production of transgenic animals, knock out animals or the analysis of naturally occurring animal models of a particular disease.

Transgenic Animals

Transgenic mice provide a useful tool for genetic and developmental biology studies and for the determination of a function of a novel sequence. According to the method of conventional transgenesis, additional copies of normal or modified genes are injected into the male pronucleus of the zygote and become integrated into the genomic DNA of the recipient mouse. The transgene is transmitted in a Mendelian manner in established transgenic strains.

Constructs useful for creating transgenic animals comprise genes under the control of either their normal promoters or an inducible promoter, reporter genes under the control of promoters to be analyzed with respect to their patterns of tissue expression and regulation, and constructs containing dominant mutations, mutant promoters, and artificial fusion genes to be studied with regard to their specific developmental outcome. Transgenic mice are useful according to the invention for analysis of the dominant effects of overexpressing a candidate gene in mouse. Typically, DNA fragments on the order of 10 kilobases or less are used to construct a transgenic animal (Reeves, 1998, New. Anat., 253:19). Transgenic animals can be created with a construct comprising a candidate gene containing one or more polymorphisms according to the invention. Alternatively, a transgenic animal expressing a candidate gene containing a single polymorphism can be crossed to a second transgenic animal expressing a candidate gene containing a different polymorphism and the combined effects of the two polymorphisms can be studied in the offspring animals. Transgenic mice engineered to overexpress a number of genes, including PCK1 (Valera et al., 1994, Proc. Natl. Acad. Sci. USA, 91: 9151), INS (Mitanchez et al., FEBS Letters, 421: 285), IAPP (D'Alession et al., 1994, Diabetes, 43:1457), Asp (Klebig et al., Proc. Natl. Acad. Sci. USA, 92: 4728) and Agrt (Graham et al., Nature Genetics,

17:273), have been prepared and may be useful for studying osteoarthritis.

Knock Out Animals

i. Standard

Knock out animals are produced by the method of creating gene deletions with homologous recombination. This technique is based on the development of embryonic stem (ES) cells that are derived from embryos, are maintained in culture and have the capacity to participate in the development of every tissue in the mouse when introduced into a host blastocyst. A knock out animal is produced by directing homologous recombination to a specific target gene in the ES cells, thereby producing a null allele of the gene. The potential phenotypic consequences of this null allele (either in heterozygous or homozygous offspring) can be analyzed (Reeves, supra). Single or double knock out mice that may be useful for studying osteoarthritis have been produced for a number of genes including IRS 1 (Araki et al., 1994, Nature, 372:186, Tamemoto et al., 1994, Nature, 372:182), 1R52 (Withers et al., 1998, Nature, 391:900), INSR, BIRKO, MIRKO, INSR (Lamothe et al., 1998, FEBS Letter, 426:381), GLUT2, GLUT4 (Katz et al., 1995, Nature, 377:151), GLP1R (Gallwitz and Schmidt, 1997, Z. Gastroenterol, 35:655):, GCK (Sakura et al., 1998, Diabetologia, 41:654), GCK/IRS1, IRS1/INSR, MC4R (Huszar et al., 1997, Cell, 88:13 1) and BRS3 (Ohki-Hamazaki et al., 1997, Nature, 390:165).

ii. *In vivo* Tissue Specific Knock Out in Mice Using Cre-lox.

The method of targeted homologous recombination has been improved by the development of a system for site-specific recombination based on the bacteriophage P1 site specific recombinase Cre. The Cre-loxP site-specific DNA recombinase from bacteriophage P1 is used in transgenic mouse assays in order to create gene knockouts restricted to defined tissues or developmental stages. Regionally restricted genetic deletion, as opposed to global gene knockout, has the advantage that a phenotype can be attributed to a particular cell/tissue (Marth, 1996, Clin. Invest. 97: 1999). In the Cre-loxP system one transgenic mouse strain is engineered such that loxP sites flank one or more exons of the gene of interest. Homozygotes for this so called 'foxed gene' are crossed with a second transgenic mouse that expresses the Cre gene under control of a cell/tissue type transcriptional promoter. Cre protein then excises DNA between loxP recognition sequences and effectively removes target gene function (Sauer, 1998, Methods, 14:381). There are now many *in vivo* examples of this method, including the inducible inactivation of mammary tissue specific genes (Wagner et al.,

1997, Nucleic Acids Res., 25:4323).

iii. Bac Rescue of Knock Out Phenotype

In order to verify that a particular genetic polymorphism/mutation is responsible for altered protein function *in vivo* one can "rescue" the altered protein function by introducing a wild-type copy of the gene in question. *In vivo* complementation with bacterial artificial chromosome (BAC) clones expressed in transgenic mice can be used for these purposes. This method has been used for the identification of the mouse circadian Clock gene (Antoch et al., 1997, Cell 89: 655).

iv. Naturally Occurring Animal Models

Naturally occurring animal models useful for studying osteoarthritis include models of severe hyperglycaemia (celebes black ape, chinese hamster, diabetes mouse (db), Djunjarian hamster, Egyptian sand rat, Hartley guinea pig, OLETF rat, New Zealand white rabbit, obese *BBZ/Wor* rat, rhesus monkey, South African hamster, spiny mouse), models for moderate hyperglycaemia (Cohen diabetic rat, GK rat, Japanese KK mouse, male Bristol CBA/Ca mouse, male eSS rat, male WKY fatty rat, male Wistar WBN/Kob rat, male ZDF rat, NZO mouse, obese mouse (ob), PBB/Ld mouse, spontaneously hypertensive corpulent (SHR/N-cp) rat, Tuco-tuco, Wellesley hybrid mouse, yellow obese mouse) and impaired glucose tolerance (ageing laboratory rats and mice, BHE rat, Fatty Zucker rat (fa), Mongolian gerbil, NON diabetic mouse, squirrel monkey, Yucatan miniature swine) (Pickup and Williams, eds., Textbook of Diabetes, 2nd Edition, Blackwell Science).

G. Production of an Amplified Product

Amplified products useful according to the invention can be prepared by utilizing the method of PCR as described in Section B entitled "Production of a Polynucleotide Sequence Primers useful for producing an amplified product according to the invention (e.g. an amplified product comprising one or more polymorphisms) can be designed and synthesized as described in Section A entitled "Design and Synthesis of Oligonucleotide Primers".

The invention provides methods (e.g. Southern blot analysis, PCR, primer extension and oligonucleotide hybridization), of detecting a polymorphism in an amplified product.

H. Production of a Mutant Protein

1. Expression of the Nucleotide Sequence

In accordance with the present invention, polynucleotide sequences which encode candidate gene protein fragments, fusion proteins or functional equivalents thereof may be used in recombinant DNA molecules that direct the expression of a candidate gene protein in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence, may be used to clone and express the candidate gene protein. As will be understood by those of skill in the art, it may be advantageous to produce candidate gene-encoding nucleotide sequences possessing non-naturally occurring codons. Codons preferred by a particular prokaryotic or eukaryotic host (Murray et al., 1989, Nucleic Acid Res 17:477) can be selected, for example, to increase the rate of protein expression or to produce recombinant RNA transcripts having desirable properties, such as a longer half-life as compared to transcripts produced from the naturally occurring sequence.

The nucleotide sequences of the present invention can be engineered in order to alter a candidate gene-encoding sequence for a variety of reasons, including but not limited to, alterations which modify the cloning, processing and/or expression of the gene product. For example, mutations may be introduced using techniques which are well known in the art, e.g., site-directed mutagenesis to insert new restriction sites, to alter glycosylation patterns, to change codon preference or to produce splice variants.

In another embodiment of the invention, a natural, modified or recombinant candidate gene protein-encoding sequence may be ligated to a heterologous sequence to encode a fusion protein (as described in Section B entitled “Production of a Polynucleotide Sequence”). For example, for screening of peptide libraries for inhibitors of candidate gene protein activity, it may be useful to encode a chimeric protein that is recognized by a commercially available antibody. a fusion protein may also be engineered to contain a cleavage site located between a candidate protein and the heterologous protein sequence, so that the protein of interest may be substantially purified away from the heterologous moiety following cleavage.

In another embodiment of the invention, the sequence encoding the candidate gene protein may be synthesized, whole or in part, using chemical methods well known in the art (see Caruthers, et al., 1980, Nuc Acids Res Symp Ser, 7:215, Horn, et al., 1980, Nuc Acids Res Symp Ser, 225, etc.) Alternatively, the protein itself, or a portion thereof, could be produced using chemical methods of synthesis. For example, peptide synthesis can be performed using various solid-phase techniques (Roberge, et al., 1995, Science, 269:202) and automated synthesis may be achieved, for example, using the A.I. 431 A Peptide Synthesizer (Perkin Elmer) in accordance with the instructions provided by the manufacturer.

The newly synthesized peptide can be substantially purified by preparative high performance liquid chromatography (e.g., Creighton, 1983, Proteins, Structures and Molecular Principles, WH Freeman and Co. New York NY). The composition of the synthetic peptides may be confirmed by amino acid analysis or sequencing (e.g., the Edman degradation procedure; Creighton, supra). Additionally the amino acid sequence of interest, or any part thereof, may be altered during direct synthesis and/or combined using chemical methods with sequences from other proteins , or any part thereof, to produce a variant polypeptide.

2. Expression Systems

In order to express a biologically active protein, the nucleotide sequence encoding the protein of interest or its functional equivalent, is inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence.

Methods which are well known to those skilled in the art can be used to construct expression vectors containing a protein-encoding sequence and appropriate transcriptional or translational controls. These methods include *in vivo* recombination or genetic recombination. Such techniques are described in Ausubel et al., supra and Sambrook et al., supra.

A variety of expression vector/host systems may be utilized to contain and express a protein product of a candidate gene according to the invention. These include but are not limited to microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (e.g., baculovirus); plant cell systems transfected with virus expression vector (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with bacterial expression vectors (e.g., Ti or pBR322 plasmid); or animal cell systems.

The “control elements” or “regulatory sequences” of these systems vary in their strength and specificities and are those nontranslated regions of the vector, enhancers, promoters, and 3’ untranslated regions, which interact with host cellular proteins to carry out transcription and translation. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid lacZ promoter of the Bluescript® phagemid (Stratagene, LaJolla CA) or pSportl (Gibco BRL) and ptrp-lac hybrids and the like may be used. The baculovirus polyhedron promoter may be used in insect cells. Promoters or enhancers derived from the genomes of plant cells (e.g., heat shock, RUBISCO; and storage protein genes) or

from plant virus (e.g. viral promoters or leader sequences) may be cloned into the vector. In mammalian cell systems promoters from the mammalian genes or from mammalian viruses are most appropriate. If it is necessary to generate a cell line that contains multiple copies of the sequence encoding the protein product of the gene of interest, vectors based on 5V40 or EBV may be used with an appropriate selectable marker.

In bacterial systems, a number of expression vectors may be selected depending upon the use intended for the protein of interest. For example, when large quantities of a protein are required for the production of antibodies, vectors which direct high level expression of fusion proteins that are readily purified may be desirable. Such vectors include, but are not limited to, the multifunctional *E. coli* cloning and expression vectors such as Bluescript® (Stratagene), in which the sequence encoding the protein of interest may be ligated into the vector in frame with sequences encoding the amino-terminal Met and the subsequent 27 residues of b-galactosidase so that a hybrid protein is produced; pIN vectors (Van Heeke & Schuster, 1989, J Biol Chem 264:5503); and the like. Pgex vectors (Promega, Madison WI) may also be used to express foreign polypeptides as fusion proteins with GST. In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems are designed to include heparmn, thrombin or factor XA protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will.

In the yeast, Saccharomyces cerevisiae, a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase and PGH may be used. For reviews, see Ausubel et al (supra) and Grant et al., 1987, Methods in Enzymology 153:516.

In cases where plant expression vectors are used, the expression of a sequence encoding a protein of interest may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV (Brisson et al., 1984, Nature 310:511) may be used alone or in combination with the omega leader sequence from TMV (Takamatsu et al., 1987, EMBO J 6:307). Alternatively, plant promoters such as the small subunit of RUBISCO (Coruzzi et al., 1984, EMBO J 3:1671; Broglie et al., 1984, Science, 224:838); or heat shock promoters (Winter I and Sinibaldi RM, 1991, Results Probl Cell Differ., 17:85) may be used. These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transection. For reviews of such techniques, see Hobbs S or Murry LE in McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill New York NY, pp 191-196 or Weissbach and Weissbach (1988) Methods for Plant Molecular Biology, Academic Press, New York, pp 421-463.

An alternative expression system which could be used to express a protein of interest is an

insect system. In one such system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in *Spodoptera frugiperda* cells or in *Trichoplusia* larvae. The sequence encoding the protein of interest may be cloned into a nonessential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of the sequence encoding the protein of interest will render the polyhedron gene inactive and produce recombinant virus lacking coat protein coat. The recombinant viruses are then used to infect *S. frigoerda* cells or *Trichoplusia* larvae in which the protein of interest is expressed (Smith et al., 1983., J Virol 46:584; Engelhard, et al., 1994, Proc Natl Acad Sci 91:3224).

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, a sequence encoding the protein of interest may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a nonessential El or E3 region of the viral genome will result in a viable virus capable of expressing in infected host cells (Logan and Shenk, 1984, Proc Natl Acad Sci, 81:3655). In addition, transcription enhancers, such as the rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

Specific initiation signals may also be required for efficient translation of a sequence encoding the protein of interest. These signals include the ATG initiation codon and adjacent sequences. In cases where the sequence encoding the protein, its initiation codon and upstream sequences are inserted into the most appropriate expression vector, no additional translational control signals may be needed. However, in cases where only coding sequence, or a portion thereof, is inserted, exogenous transcriptional control signals including the ATG initiation codon must be provided. Furthermore, the initiation codon must be in the correct reading frame to ensure transcription of the entire insert. Exogenous transcriptional elements and initiation codons can be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate to the cell system in use (Scharf, et al., 1994, Results Probl Cell Differ, 20:125; Bittner et al., 1987, Methods in Enzymol, 153:516).

In addition, a host cell strain may be chosen for its ability to modulate the expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be important for correct insertion, folding and/or function. Different host cells such as CHO, HeLa, MDCK, 293, W138, etc have specific cellular machinery and characteristic mechanisms for such post-translational activities and may be chosen to ensure the correct modification and processing

of the introduced, foreign protein.

For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines which stably express a foreign protein may be transformed using expression vectors which contain viral origins of replication or endogenous expression elements and a selectable marker gene. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clumps of stably transformed cells can be expanded using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase (Wigler., et al., 1977, Cell 11:223) and adenine phosphoribosyltransferase (Lowy, et al., 1980, Cell 22:817) genes which can be employed in tk- or aprt- cells, respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dhfr which confers resistance to methotrexate (Wigler et al., 1980, Proc Natl Acad Sci 77:3567); npt, which confers resistance to the aminoglycosides neomycin and G-418 (Colbere-Garapin et al., 1981., J Mol Biol., 150:1) and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, supra). Additional selectable genes have been described, for example, trpB, which allows cells to utilize indole in place of tryptophan, or hisD, which allows cells to utilize histinol in place of histidine (Hartman and Mulligan, 1988, Proc Natl Acad Sci 85:8047). Recently, the use of visible markers has gained popularity with such markers as anthocyanins, B glucuronidase and its substrate, GUS, and luciferase and its substrate, luciferin, being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes et al., 1995, Methods Mol Biol 55:121).

3. Identification of Transformants Containing the Polynucleotide Sequence

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, its presence and expression should be confirmed. For example, if the sequence encoding a foreign protein is inserted within a marker gene sequence, recombinant cells containing the sequence encoding the foreign protein can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with the sequence encoding the foreign protein under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem sequences as well.

Alternatively, host cells which contain the coding sequence for a protein of interest and express the protein of interest may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridization and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of the nucleic acid or protein.

The presence of the polynucleotide sequence encoding the protein of interest can be detected by DNA-DNA or DNA-RNA hybridization or amplification using probes, portions or fragments of the sequence encoding the foreign protein of interest.

A variety of protocols for detecting and measuring the expression of the foreign protein, using either polyclonal or monoclonal antibodies specific for the protein are known in the art. Examples include enzyme-linked immunosorbant assay (ELISA), radioimmunoassay (RIA) and fluorescent activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on the protein of interest is preferred, but a competitive binding assay may be employed. These and other assays are described in Hampton et al., 1990, Serological Methods a Lahoratory Manual, APS Presds, St Paul MN and Maddox., et al., 1983, J Exp Med 158:1211.

4. Purification of the Protein of Interest

Host cells transformed with a nucleotide sequence encoding a protein of interest may be cultured under conditions suitable for the expression and recovery of the encoded protein from cell culture. The protein produced by a recombinant cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing a sequence encoding a protein of interest can be designed with signal sequences which direct secretion of the protein of interest through a prokaryotic or eucaryotic cell membrane. Other recombinant constructions may join the sequence encoding the protein of interest to the nucleotide sequence encoding a polypeptide domain which will facilitate purification of soluble proteins (Kroll et al., 1993, DNA Cell Biol, 12:441).

The protein of interest may also be expressed as a recombinant protein with one or more additional polypeptide domains added to facilitate protein purification. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as a histidine-tryptophan modules that allow purification on immobilized metals, protein a domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp, Seattle WA). The inclusion of a cleavable linker sequences such as Factor XA or enterokinase

(Invitrogen, San Diego CA), between the purification domain and the protein of interest is useful for facilitating purification. One such expression vector provides for expression of a fusion protein comprising the sequence encoding a foreign protein and nucleic acid sequence encoding 6 histidine residues followed by thioredoxin and an enterokinase cleavage site. The histidine residues facilitate purification while the enterokinase cleavage site provides a means for purifying the foreign protein from the fusion protein.

In addition to recombinant production, fragments of the protein of interest may be produced by direct peptide synthesis using solid-phase techniques (Stewart et al., 1969, Solid-Phase Peptide Synthesis, WH Freeman Co,. San Francisco; Merrifield, 1963, J Am Chem Soc, 85:2149). *In vitro* protein synthesis may be performed using manual techniques or by automation. Automated synthesis may be achieved, for example, using Applied Biosystems 431 A Peptide Synthesizer (Perkin Elmer, Foster City CA) in accordance with the instructions provided by the manufacturer. Various fragments of a protein of interest may be chemically synthesized separately and combined using chemical methods to produce the full length molecule.

I. Preparation of Antibodies

Antibodies specific for the protein products of the candidate genes of the invention are useful for protein purification, for the diagnosis and treatment of various diseases (e.g osteoarthritis) and for drug screening and drug design methods useful for identifying and developing compounds to be used in the treatment of various diseases (e.g. osteoarthritis). By antibody, we include constructions using the binding (variable) region of such an antibody, and other antibody modifications. Thus, an antibody useful in the invention may comprise a whole antibody, an antibody fragment, a polyfunctional antibody aggregate, or in general a substance comprising one or more specific binding sites from an antibody. The antibody fragment may be a fragment such as an Fv, Fab or $F(ab')_2$ fragment or a derivative thereof, such as a single chain Fv fragment. The antibody or antibody fragment may be non-recombinant, recombinant or humanized. The antibody may be of an immunoglobulin isotype, e.g., IgG, 1gM, and so forth. In addition, an aggregate, polymer, derivative and conjugate of an immunoglobulin or a fragment thereof can be used where appropriate. Neutralizing antibodies are especially useful according to the invention for diagnostics, therapeutics and methods of drug screening and drug design.

Although a protein product (or fragment or oligopeptide thereof) of a candidate gene of the invention that is useful for the production of antibodies does not require biological activity, it must be antigenic. Peptides used to induce specific antibodies may have an amino acid sequence consisting of

at least five amino acids and preferably at least 10 amino acids. Preferably, they should be identical to a region of the natural protein and may contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of amino acids corresponding to the protein product of a candidate gene of the invention may be fused with amino acids from another protein such as keyhole limpet hemocyanin or GST, and antibody will be produced against the chimeric molecule. Procedures well known in the art can be used for the production of antibodies to the protein products of the candidate genes of the invention.

For the production of antibodies, various hosts including goats, rabbits, rats, mice etc... may be immunized by injection with the protein products (or any portion, fragment, or oligonucleotide thereof which retains immunogenic properties) of the candidate genes of the invention. Depending on the host species, various adjuvants may be used to increase the immunological response. Such adjuvants include but are not limited to Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, and dinitrophenol. BCG (bacilli Calmette-Guerin) and *Corynebacterium parvum* are potentially useful human adjuvants.

1. Polyclonal antibodies.

The antigen protein may be conjugated to a conventional carrier in order to increase its immunogenicity, and an antiserum to the peptide-carrier conjugate will be raised. Coupling of a peptide to a carrier protein and immunizations may be performed as described (Dymecki et al., 1992, J. Biol. Chem., 267: 4815). The serum can be titered against protein antigen by ELISA (below) or alternatively by dot or spot blotting (Boersma and Van Leeuwen, 1994, J Neurosci. Methods, 51: 317). At the same time, the antiserum may be used in tissue sections prepared asdescribed. A useful serum will react strongly with the appropriate peptides by ELISA, for example, following the procedures of Green et al., 1982, Cell, 28: 477.

2. Monoclonal antibodies.

Techniques for preparing monoclonal antibodies are well known, and monoclonal antibodies may be prepared using a candidate antigen whose level is to be measured or which is to be either inactivated or affinity-purified, preferably bound to a carrier, as described by Arnheiter et al., 1981, Nature, 294;278.

Monoclonal antibodies are typically obtained from hybridoma tissue cultures or from ascites

fluid obtained from animals into which the hybridoma tissue was introduced.

Monoclonal antibody-producing hybridomas (or polyclonal sera) can be screened for antibody binding to the target protein.

3. Antibody Detection Methods

Particularly preferred immunological tests rely on the use of either monoclonal or polyclonal antibodies and include enzyme-linked immunoassays (ELISA), immunoblotting and immunoprecipitation (see Voller, 1978, Diagnostic Horizons, 2:1, Microbiological Associates Quarterly Publication, Walkersville, MD; Voller et al., 1978, J. Clin. Pathol., 31: 507; U.S. Reissue Pat. No. 31,006; UK Patent 2,019,408; Butler, 1981, Methods Enzymol., 73: 482; Maggio, E. (ed.), 1980, Enzyme Immunoassay, CRC Press, Boca Raton, FL) or radioimmunoassays (RIA) (Weintraub, B., Principles of radioimmunoassays, Seventh Training Course on Radioligand Assay Techniques, The Endocrine Society, March 1986, pp. 1-5, 46-49 and 68-78). For analysing tissues for the presence or absence of a protein produced by a candidate gene according to the present invention, immunohistochemistry techniques may be used. It will be apparent to one skilled in the art that the antibody molecule may have to be labelled to facilitate easy detection of a target protein. Techniques for labelling antibody molecules are well known to those skilled in the art (see Harlow and Lane, 1989, Antibodies, Cold Spring Harbor Laboratory).

J. Preparation of a Labeled Protein

1. Labling of protein

Labeling techniques are useful, according to the invention, for studying the biochemical properties, processing, intracellular transport, secretion and degradation of proteins.

Biosynthetic labeling of proteins produced by candidate genes of the invention is preferably performed with ^{35}S-methionine due to the high specific activity (>800Ci/mmol) and ease of detection of this amino acid. Another amino acid should be used to label a protein that contains little or no methionine.

According to the following protocol, either suspension cells or adherent cells are labeled with ^{35}S-methionine. Briefly, cells are washed and incubated for 15 min at 37^{0}C in short-term labeling medium (complete serum-free, methionine free RPMI or DMEM containing 5% (v/v) dialyzed fetal bovine serum) to deplete intracellular pools of methionine. Cells are then incubated in the presence of ^{35}S-methionine working solution (0.1 to 0.2 mCi/ml in 37^{0}C short-term labeling medium) such that 4ml of ^{35}S-methionine working solution is added per 2 x 10^7 suspension cells and 2 to 4 ml of ^{35}S-

methionine working solution is added per 100 mm dish of adherent cells (0.5-2 x 10^7 cells), for a period of 30 min to 3 hour in a humidified, 37°C, 5% CO_2 incubator. Upon completion of labeling, suspension cells are washed by centrifugation in ice-cold PBS. Following removal of labeling medium, adherent cells are washed with PBS, scraped and collected by centrifugation. Labeled cells are processed and analyzed by immunoaffinity chromatography, immunoprecipitation and one- and two-dimensional gel electrophoresis (Ausubel et al., supra).

If the protein of interest is synthesized at a relatively low rate or is in a steady state, it may be necessary to label cells for an extended period of time. When performing long-term biosynthetic labeling of cells, it is necessary to include unlabeled methionine in the medium to maintain cell viability and to ensure that incorporation of label is maintained during the course of the experiment. According to this method, cells can be labeled in the presence of ^{35}S-methionine in long term labeling medium (90% methionine free RPMI or DMEM) for up to 16 hours (Ausubel et al., supra).

2. *In vitro* Translation

The protein product of the cloned candidate gene of the invention can be produced by the methods of *in vitro* transcription and *in vitro* translation. *In vitro* transcription is performed essentially as described in Section B entitled "Production of a Polynucleotide Sequence" in the absence of a labeled ribonucleoside. The RNA produced by the *in vitro* transcription reaction will be extracted with phenol, ethanol precipitated twice and resuspended in 10ml of TE buffer. *In vitro* translation is performed by adding 1 to 10ml of RNA to an *in vitro* translation kit (e.g. wheat germ or reticulocyte lysate) in the presence of 15mCi [^{35}S]methionine, following the directions provided by the manufacturer. A typical reaction is carried out in a 30ml volume at room temperature for 30 to 60 minutes (Ausubel et al., supra).

K. Production of Cells Expressing a Nucleotide Sequence Comprising a Polymorphism

Mammalian cells expressing a nucleotide sequence comprising a polymorphism are useful, according to the invention for determining the biochemical and functional properties of the protein product of a nucleotide sequence comprising a polymorphism, for analyzing expression of a candidate gene, for large scale production of a protein of interest, for drug screening and for the production of transgenic animals or knockout mice.

Methods of efficiently introducing foreign DNA into mammalian cells are known in the art and include calcium phosphate transfection, DEAE-dextran transfection, electroporation and liposome-

mediated transfection (Ausubel et al., supra).

Transfection Protocols

1. Calcium-Phosphate Transfection

The method of calcium phosphate transfection involves preparing a precipitate by slowly mixing a HEPES-buffered saline solution with a mixture of calcium chloride and DNA. According to this method, up to 10% of the cells on a dish will incorporate DNA.

Cells to be transfected are split one day prior to transfection so that on the day of transfection cells are well-separated on the plate. a 10 cm dish of cells is fed with 9.0 ml of complete medium approximately 2 to 4 hours before the addition of the precipitate. DNA to be transfected (10-50mg/10-cm plate) is ethanol precipitated, resuspended in 450 ml sterile water and mixed with 50 ml of 2.5 M $CaCl_2$. The DNA/$CaCl_2$ solution is added dropwise to a 15-ml conical tube containing 500 ml 2X HeBS (0.283M NaCl, 0.023M HEPES acid, 1.5 mM Na_2HPO_4, pH 7.05). It is preferable to bubble the HeBS solution during the addition of the DNA mixture. After the precipitate has formed for 20 minutes at room temperature, it is added evenly to the cells. The cells are incubated with the precipitate at 37°C in a CO_2 humidified incubator for 4-16 hours. Following removal of the precipitate, the cells are washed with PBS and fed in complete medium. Glycerol or dimethyl sulfoxide shock can be used to increase the DNA uptake by certain types of cells (Ausubel et al., supra).

2. DEAE-Dextran Transfection

Cells to be transfected are plated at a concentration such that after 3 days of growth they are 30-50% confluent. The DNA to be transfected (approximately 4 mg) is ethanol precipitated, resuspended in 40ml TBS and added slowly while shaking to 80 ml of warm 10 mg/ml DEAE-dextran in TBS. After cells have been washed with PBS and fed with 4 ml of DMEM containing 10% Nu Serum/10cm dish, the DEAE-dextranlDNA mixture is evenly distributed over the entire plate. Cells are incubated with the DNA for approximately 4 hours in a humidified CO_2 incubator. Following the removal of the DEAE-dextran/DNA mixture, cells are shocked by the addition of 5 ml of 10% DMSO in PBS. After a 1 minute incubation at room temperature, cells are washed with PBS and fed with complete medium (Ausubel et al., supra).

3. Electroporation

Alternatively, DNA can be introduced into cells by the use of high-voltage electric shocks, a technique termed electroporation. Briefly, according to the method of electroporation, cells are

suspended in an appropriate electroporation buffer and placed in an electroporation cuvette. Following the addition of DNA, the cuvette is connected to a power supply and the cells are subjected to a high-voltage electrical pulse of a defined magnitude and length, optimized for the cell type being transfected. After a brief period of recovery, the cells are placed in normal culture medium.

A population of cells to be transfected by electroporation is grown to late-log phase in complete medium. Typically stable transfection requires 5 X 106 cells, and transient transfection requires 1-4 X 10^7 cells. Cells are harvested by centrifugation for 5 minutes at 640 x g at 4^0C. The resulting cell pellet is resuspended in half of the original volume of ice-cold electroporation buffer (e.g. PBS without calcium or magnesium, Hepes buffered saline, tissue culture medium without serum, or phosphate buffered sucrose (272mM sucrose/7 mM K_2HPO_4, pH 7.4/1mM $MgCl_2$)). The choice of an electroporation buffer is dictated by the cell line. Cells are then harvested by centrifugation for 5 minutes at 640 x g at 4^0C, and resuspended at 1 X 10^7/ml in electroporation buffer at 0^0C for stable transfection or at a higher concentration (up to 8 X 10^7/ml) for transient transfection. Aliquots of the cells (0.5 ml) are transferred into the desired number of electroporation cuvettes and placed on ice.

DNA is added to the cell suspension in the cuvettes on ice. For stable transfection, DNA (optimally 1-10 mg) should be linearized with a restriction enzyme that cuts at a site in a non-essential region, purified by phenol extraction and ethanol precipitated. Supercoiled DNA (optimally 10 mg) may be used for transient transfection. The DNA/cell suspension is mixed, and incubated on ice for 5 minutes.

The cuvette is placed in the holder in the electroporation apparatus (at room temperature) and shocked one or more times at the desired voltage and capacitance settings. An electroporation apparatus useful according to the invention is the Bio-Rad Gene Pulser. The number of shocks and the voltage and capacitance settings will vary depending on the cell type, and should be optimized. The two parameters that are critical for successful electroporation are the maximum voltage for the shock and the duration of the current pulse.

Following electroporation, the cuvette containing the mixture of cells and DNA is incubated on ice for 10 minutes. The transfected cells are diluted 20-fold in complete culture medium. For stable transfection cells are grown for 48 hours in nonselective medium and then transferred to antibiotic containing medium. For transient transfection, cells are incubated 50-60 hours and then harvested for the desired transient assay.

L. Production of Animals Expressing a Nucleotide Sequence Comprising a Polymorphism

Transgenic animals expressing a construct comprising a candidate gene containing a polymorphism, according to the invention can be produced by methods well known in the art (reviewed in Reeves et al., supra). Knock out mice wherein a candidate gene according to the invention has been disrupted can be produced by methods well known in the art (reviewed in Moreadith and Radford, 1997, J,Mol. Med., 75:208 and Shastry, 1998, Mol. Cell. Biochem., 181:163). These animals provide useful models for studying the functional consequences of one or more polymorphisms in a gene of interest.

M. Production of a Candidate Gene Library

The invention provides a method of producing a candidate gene library comprising genes that are potentially associated with the susceptibility to, or pathogenesis of a disease. A candidate gene library is useful for determining the genetic basis of a disease of interest.

Genetic susceptibility to a disease must occur as a result of specific DNA differences relative to non-susceptible individuals. In the case of osteoarthritis, many genes are known which are potentially involved in the susceptibility to, or pathogenesis of the disease. These genes are included in the candidate gene library and the association of these genes with osteoarthritis is determined from population studies according to the invention. Unlike linkage studies wherein a region of the genome that is thought to be involved in a disease is determined, the candidate gene strategy, including association studies, addresses the involvement of a particular gene in a disease. The results of association studies of candidate genes are used to identify genes that should be intensively studied as potential therapeutics or therapeutic targets.

According to the invention, the full range of polymorphic sites within each candidate gene is identified and examined in diseased and normal populations. The frequency of each gene variant (allele) in each population is then compared to the other. If a specific polymorphism under analysis contributes to the disease phenotype, it will be present in the diseased population at a higher frequency than in the normal population. In addition, if the specific polymorphism under analysis does not itself contribute to the disease phenotype but resides elsewhere in, or is near to a gene containing a contributory polymorphism, a significant association may be seen with the polymorphic marker being tested. This is because the two markers are in linkage disequilibrium with each other due to their close proximity.

1. Strategies for Identifying Genes Associated with a Disease

There are a number of methods known in the art for the identification of genes involved in a

disease. These methods include familial linkage studies followed by positional cloning, differential gene expression studies on tissues, and population-based candidate gene association studies. Although positional cloning has proven to be useful for diseases resulting from a single mutation, this technique is not suitable for identifying genetic linkage in diseases where multiple genetic variants combine to create disease susceptibility. Furthermore, it has been demonstrated that the etiological basis of the majority of diseases comprises more than one gene.

The goal of linkage studies is to determine the approximate position of disease genes by studying related individuals in families. According to linkage strategies, DNA markers that are randomly spaced throughout the genome, but are rarely located within genes, are tested for the frequency of their presence along with the particular disease phenotype. There is approximately a 50% chance of an unlinked gene and marker gene co-localizing. If a particular marker is present at a significantly higher frequency than expected in disease individuals, this indicates that the marker is located in the vicinity of the disease gene. Usually the disease gene is delimited to a large region (containing tens to hundreds of genes). After a disease gene has been grossly mapped, this entire region must be extensively characterized to determine what genes are present in the region. Any gene that is identified according to this method becomes a candidate gene.

Linkage studies have been used successfully to identify the genes responsible for certain genetic diseases originating from mutations in a single gene (monogenic diseases). However, most common human diseases are of polygenic origin wherein changes in multiple genes causes an increased susceptibility to or pathogenesis of a particular disease. Because the DNA changes associated with genes which contribute to polygenic diseases are common in the population, thereby diluting the contribution of a given region of the genome to the disease, it is difficult to perform linkage studies on diseases of polygenic origin.

Linkage analysis

A series of genetic crosses is performed in an animal model system of a particular defect that is characteristic of a disease of interest (e.g. osteoarthritis) between individuals having an observable mutant phenotype and normal individuals of a control strain. At least one disease- related loci is used as a marker in these crosses. Alternatively, linkage analysis can be performed using chromosomal markers that do not comprise a disease related locus (described below). If non-random assortment of the mutant trait with a marker locus is observed, and if that non-random assortment is statistically significant (for example, if a Student's t test or ANOVA is applied to the results) the trait is linked to the marker locus.

Similarly, linkage analysis using an existing human or other mammalian pedigree may be performed. Pedigree analysis is a useful technique for identifying genes for which variant alleles may contribute to the risk, onset or progression of a disease in a family containing multiple individuals afflicted with a disease; according to this method, numerous genetic loci from affected and unaffected family members are compared. Non-random assortment of a given genetic marker between affected and unaffected family members relative to the distributions observed for other genetic loci indicates that the marker (for example, a variant isoform of a gene) either contributes to the disease or is in physical proximity to another that does so.

If a non-random assortment of the disease-related phenotype with a marker locus is observed, using either approach, this is indicative of an association between the gene underlying the defect and that locus. Because the strength of any conclusion drawn from linkage analysis is statistically-based, the accuracy of the results is thought to be proportional to the number of crosses or family members and genetic loci analyzed.

Positional Cloning

If linkage is confirmed it is preferable to perform a molecular analysis of the region in which the peak of linkage maps. The wide availability of yeast artificial chromosome (YAC) or bacterial artificial chromosome (BAC) libraries facilitates this analysis. a nucleic acid sequence specific for a region encompassing a gene which is determined to occupy a map location of a particular locus of interest is examined, and open reading frames are evaluated to determine their relationship with the observed phenotype. An initial evaluation may be performed with the assistance of a computer program, such as the PathCalling™ (CuraGen) biological pathway discovery platform. All or a subset of the open reading frames present in the region are then cloned (e.g., by PCR) from mutant animals or affected family members and from their healthy counterparts (either control animals or unaffected family members), and the sequences of these open reading frames are compared. If a mutation or other allelic variant is found to be linked to individuals displaying the disease phenotype (in a statistically-significant, non-random manner), it can be concluded that this mutation is associated with a disease phenotype. A nucleic acid fragment containing this gene can be labeled and used as a probe for *in situ* hybridization analysis of fixed chromosomes of the human or other mammal to determine precisely the physical location of the gene. Furthermore, a gene that has been mapped and isolated in this manner may be useful as a candidate target for disease diagnosis and for drug targeting according to the invention (see below).

2. Identification of Genes to be Included in Candidate Gene Library

A candidate gene library according to the invention will include i. genes that are involved in known or predicted disease pathways, ii. new genes that are identified by a relevant pattern of specific tissue or cell expression, iii. genes that map to genomic regions of known linkage, and iv. gene sequences (from sequence databases) that are homologs of the above referenced categories of potential candidate genes. The choice of potentially related genes to be selected from a database will depend on the percent identity as calculated by Fast DB and based upon mismatch penalty, gap penalty, gap size penalty and joining penalty. Figure 1 summarized

Based on the physiological changes associated with a disease of interest, predictions can be made regarding a cell or tissue-type that would be expected to express high or low levels of candidate genes associated with a particular disease. For osteoarthritis, it is expected that muscle, adipose, pancreas or liver tissue or tissue comprising insulin secreting pancreatic b-cells, would be useful for identifying candidate genes according to the invention.

Differences in the expression of known and unknown genes in normal and disease tissue can be determined by methods known in the art including Serial Analysis of Gene Expression (SAGE) (Velcuescu et al., 1995, Science, 270:484), subtractive hybridization/screening (described below), differential display (Ling and Pardee, 1992, Science, 257:967) high-density microarray expression testing.

The technique of SAGE allows for the rapid, detailed analysis of thousands of transcripts. SAGE depends on the following two principles. First, sufficient information is contained within a short nucleotide sequence (approximately 9-lObp), isolated from a defined location within a transcript, to uniquely identify a transcript. Second, the concatenation of short tags of sequence allows transcripts to be analyzed serially by sequencing multiple tags within a single clone.

The method of SAGE is performed by synthesizing double-stranded cDNA from mRNA, cleaving the resulting cDNA with an anchoring restriction endonuclease that is expected to cleave most transcripts at least one time, and isolating the most 3' region of the cleaved cDNA by binding to streptavadin beads. This protocol allows for the identification of a unique site on a transcript that corresponds to the restriction site located closest to the polyA tail. Replicate samples of the most 3' region of the cDNA are ligated to one of two linker molecules that contain a type IIS restriction site for a tagging enzyme. The cleavage site for Type IIS restriction endonucleases is located at a defined distance up to 20 bp from the asymmetric recognition site. Linkers are designed such that upon cleavage of the ligation product with the tagging enzyme there is release of the linker and an attached short region of cDNA.

Following the creation of blunt ends, the two pools of released tags are ligated to each other and the resulting ligated product is used as a template for PCR amplification in the presence of primers that are specific for each linker. The PCR product is cleaved with the anchoring enzyme and amplification products, comprising two tags linked tail to tail, are isolated, concatenated by ligation, cloned and sequenced (Velescu et al., supra).

Differential display provides a method for separating and cloning individual mRNAs by PCR analysis. According to the method of differential display, oligonucleotide primers are selected wherein one primer is anchored to the polyadenylate tail of a subset of mRNA species and the other primer is short and of an arbitrary sequence such that it anneals at different positions relative to the first primer. The mRNA subpopulations that are identified with these primer pairs are subjected to reverse transcription, amplified and analyzed on a DNA sequencing gel. By using multiple sets of primers, a reproducible pattern of amplified cDNA fragments that demonstrate a requirement for the sequence specificity of either primer can be obtained (Liang and Pardee, supra).

According to the method of high-density microarray expression testing, DNA sequences to be tested for expression are spotted onto a surface, usually at high-density to allow for the testing of many genes. The surface contain the DNA sequences is typically referred to as a 'chip'. The spotted DNA cam be either cDNA clones or oligonucleotides. RNA is prepared from the two cells or tissues to be compared. The RNA from one cell/tissue will be labeled red and the RNA from the other cell/tissue will be labeled yellow. Both RNA preparations are hybridized to the DNA array. The ratio of red to yellow is indicative of the relative levels of expression between the two cells/tissues.

3. Mapping a candidate gene

Molecular and cytogenetic methods of mapping candidate genes are known in the art and are summarized below. Linkage analysis provides a method for identifying genes mapping to genomic regions of known linkage.

<u>Linkage analysis</u>

As described above, linkage analysis may be performed between an unmapped candidate gene and one or more of the disease-related loci or by analyzing the genetic linkage between the candidate gene and chromosomal markers which are not themselves linked to a disease-related locus, according to the same method. For the latter type of analysis it is preferable that the spacing of markers throughout the genome of the test organism is approximately one every cM or less. This spacing will ensure complete coverage of the genome and will facilitate accurate mapping.

Other methods for mapping a candidate gene are provided below.

Syntenic similarity

As a result of classical genetic studies and, more recently, multi-laboratory genomic sequencing collaborations such as the Human Genome Project and Mouse Genome Project, the human and mouse genomes have been extensively characterized. It is now known that there is a significant degree of co-linearity among human, mice and rats wherein there is conservation relative to one another among these several species in the chromosomal map positions of numerous genes and groups of genes. Examination of the human and/or mouse chromosomal maps in the regions comparable to those to which a particular loci of interest maps in the rat will yield candidate genes which may be responsible for the physiological changes associated with a disease of interest. The methods of radiation hybrid mapping or fluorescence *in situ* hybridization at low stringency to rat chromosomes using labeled fragments derived from the human or mouse genes can be used to confirm that genes present in these regions of the human and/or mouse are present in the regions of interest in the rat.

Radiation hybrid (RH) mapping is a somatic cell hybrid technique that was developed to create high resolution, contiguous maps of mammalian chromosomes. The method is useful for ordering DNA markers spanning millions of base pairs of DNA at a resolution not easily obtained by other mapping methods (Cox et al., 1990, Science, 250: 245; Burmeister et al., 1991, Genomics, 9:19; Warrington et al., 1992, Genomics, 13: 803; Abel et al., 1993, Genomics, 17:632). Radiation hybrid mapping facilitates the mapping of non-polymorphic DNA markers that cannot be used for meiotic mapping.

According to the method of radiation hybrid mapping a lethal dose of X-irradiation is used to fragment the chromosomes of the donor cell line. Chromosome fragments from the donor cell line are then retained, in a non-selective manner, following cell fusion with a recipient cell line. The resulting hybrid clones are then analyzed for the presence or absence of specific donor chromosome markers. It is expected that markers that are further apart on a chromosome are more likely to be broken apart by radiation and to segregate independently in the RH cells than markers that are closer together. By performing a statistical analysis of the co-segregation of various loci in hybrid clones, it is possible to construct a map that provides information regarding the relative order and distance of markers (Cox et al., 1990, supra; Warrington et al., 1991, Genomics, 11: 701; Ceccherini et al., 1992, Proc. Natl. Acad. Sci. USA, 89: 104).

Subtractive screening

In view of the observation that only a subset of an organism's genes are expressed in a given tissue, there is a high probability that transcripts which differ in expression between cells of the same tissue in a mutant and control animal are responsible for the observed mutant phenotype.

According to the method of subtractive cloning, mRNA is isolated from a tissue of choice, wherein the tissue is obtained from two distinct organisms and wherein one organism displays a mutant phenotype with regard to a particular trait while the other is normal in that respect. Methods well known in the art are used to prepare cDNA from the mRNA derived from the organism. The mRNA template is then degraded, either by hydrolysis under alkaline conditions or by RNAase H-mediated cleavage, and the cDNA is returned to a buffer in which mRNA is stable, and mixed with a molar excess of mRNA prepared from the second organism under conditions of stringent hybridization. The mixture is then passed over a hydroxyapatite column, which binds double-stranded nucleic acids but allows single stranded nucleic acid molecules to pass through. Reverse transcripts derived from the first sample which do not hybridize to niRNA molecules derived from the second organism (in other words, reverse transcripts specific to the first tissue sample) are present in the flow-through fraction and are cloned into a vector to create a subtraction library. The reciprocal experiment (in which the cDNA is derived from the second mRNA preparation) is also carried out to create a complete set of transcripts specific to the tissue samples derived from the two organisms.

This procedure will provide transcripts that can be labeled and used as probes in *in situ* hybridization analysis of immobilized chromosomes. The method of subtractive screening therefore, yields both cloned genes as well as reagents useful for determining if the cloned genes co-localize with a loci of interest. If a particular gene is found to co-localize to a loci of interest, the genes may be analyzed functionally (e.g., in a phenotypic rescue experiment, as described below or by the phenotypic assays described in Section F entitled "Identification and Characterization of Polymorphisms") Ultimately, these genes may be used as targets for drugs or disease diagnostic methods, or even as therapeutic nucleic acids.

Mutagenic transposon mapping

The selection of insertional events that lie within genes (e.g., within coding or regulatory sequences) is facilitated by the use of entrapment vectors, first described in bacteria (Casadaban and Cohen, 1979, Proc. Natl. Acad. Sci. U.S.A., 76: 4530; Casadaban et al., 1980, J Bacteriol, 143: 971). By employing animal models, entrapment vectors can be introduced into pluripotent ES cells in culture (for example, using electroporation or a retrovirus) and then passed into the germline via chimeras

(Gossler et al., 1989, Science, 244: 463; Skarnes, 1990, Biotechnology, 8:827). Alternatively, transgenic animals containing entrapment vectors may be generated by standard oocyte injection protocols.

These methods result in DNA integrations that are highly mutagenic because they interrupt the endogenous coding sequence. It is estimated that the frequency of obtaining a mutation in some gene of any in the genome using a promoter or gene trap is about 45%. For adetailed description of retroviral insertion mutagenesis see Methods Enzymol., vol. 225, 1990. Genes which are expressed in a tissue of interest and for which a biochemical assay of a particular activity have been developed in animal models are most useful according to this method. Promoter or gene trap vectors often contain a reporter gene, e.g., *lacZ, Cat* or *green fluorescent protein (Gfp)* that lacks its own upstream promoter and/or splice acceptor sequence. That is, promoter gene traps contain a reporter gene with a splice site but no promoter. If the vector integrates within a gene and is spliced into the gene product, then the reporter gene will be expressed. Enhancer traps contain a reporter gene and have a minimal promoter which requires the activity of an enhancer in order to function. If the vector integrates near an enhancer (whether in a gene or not), then the reporter gene will be expressed. Activation of the reporter gene can only occur when the vector is integrated within an active host gene and generates a fusion transcript with the host gene. The activity of a reporter gene provides an easy assay for determining if a vector has been integrated into an expressed gene. Methods for detecting reporter gene activity in transfected cells or tissues of a transgenic animal are well known in the art.

The mutagenic vector may be mapped using standard cytogenetic techniques, such as *in situ* hybridization, wherein a labeled fragment comprising vector-specific sequence is used as a probe. Co-localization of the probe with a particular locus of interest indicates that the associated gene is a suitable candidate and should be subjected to further analysis. A gene that has been identified in this manner can be cloned as described.

N. Diagnostic Indicators, Screens and Disease Symptoms

In another embodiment of the invention, there is provided a method of diagnosing or determining susceptibility of a subject to joint space narrowing and/or osteophyte development and/or joint pain. This method involves analyzing the genetic material of a subject to determine which allele(s) of a gene is/are present. The method may include determining whether one or more particular alleles are present, or which combination of alleles (i.e. a haplotype) is present. The method may also include determining whether subjects are homozygous or heterozygous for a particular allele or haplotype.

In a preferred embodiment, the method comprises determining which allele of one or more

polymorphisms of the invention is/are present. In particular, the method may include determining the presence of a polymorphism of a gene which in combination with polymorphisms defined herein or other polymorphisms may define a risk haplotype. The polynucleotides sequences for these particular alleles may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotides, complementary RNA and DNA molecules and PNAs. The polynucleotides may be used to determine whether subjects are homozygous or heterozygous for a particular allele or haplotype making them susceptible to joint space narrowing and/or osteophyte development and/or joint pain, and hence, osteoarthritis.

In one aspect, hybridization with a PCR probe which is capable of detecting a particular polymorphism may be used to identify nucleic acid sequences of particular alleles or haplotype. These probes must be specific to these particular alleles and the stringency of the hybridization or amplification must be such that the probe identifies only this particular allele.

Means for producing specific hybridization probes for these polynucleotides of particular alleles include the cloning of these polynucleotide sequences into vectors for the production of mRNA probes is well known to one skilled in the art. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotides of particular alleles or haplotype may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect susceptibility to joint space narrowing and/or osteophyte development and/or joint pain. Such qualitative methods are well known in the art.

In a particular embodiment, polynucleotides of particular alleles or haplotype may be used in assays that detect susceptibility to joint space narrowing and/or osteophyte development and/or joint pain, particularly those mentioned above. Polynucleotides complementary to sequences of a particular allele or haplotype may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After

a suitable incubation period, the sample is washed and it is determined if there is a signal. If a signal is found, then the presence of the polynucleotide of a particular allele, alleles or haplotype in the sample indicates the susceptibility to joint space narrowing and/or osteophyte development and/or joint pain, and hence, osteoarthritis. Such assays may also be used to determine the particular therapeutic treatment regimen for an individual patient.

With respect to osteoarthritis, the presence of a particular polymorphism or polymorphisms in a tissue sample from an individual may indicate a predisposition for joint space narrowing and/or osteophyte development and/or joint pain, or may provide a means for detecting osteoarthritis prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier, thereby preventing the development or further progression of osteoarthritis.

Additional diagnostic uses for oligonucleotides designed from the polynucleotide sequences of a particular allele or haplotype may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced *in vitro*. Oligomers will contain a fragment of a polynucleotide a particular allele, alleles or haplotype or a fragment of a polynucleotide complementary to the polynucleotide a particular allele, alleles or haplotype, and will be employed under optimized conditions for identification of a specific polymorphism, polymorphisms or haplotype. Oligomers may also be employed under very stringent conditions for detection of these particular DNA or RNA sequences.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotides described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques to detect a particular polymorphism, polymorphisms or haplotype simultaneously as described below. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

Microarrays may be prepared, used, and analyzed using methods known in the art (Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116;

Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662). Various types of microarrays are well known and thoroughly described in Schena, M., ed. (1999; DNA Microarrays: A Practical Approach, Oxford University Press, London).

In another embodiment, a method involves the use of antibodies in diagnosing or determining the susceptibility to joint space narrowing and/or osteophyte development and/or joint pain. The antibodies would specifically bind to an epitope of a particular allele or form of the protein and may be used to determine susceptibility to joint space narrowing and/or osteophyte development and/or joint pain, and hence, osteoarthritis. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above. Diagnostic assays for determining susceptibility to joint space narrowing and/or osteophyte development and/or joint pain include methods which utilize the antibody and a label to detect a particular allele or form of the protein in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules are known in the art and may be used.

A variety of protocols for measuring a particular allele or form of the protein, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing susceptibility to joint space narrowing and/or osteophyte development and/or joint pain.

O. Preparation of a Human Sample

The presence of an allelic form of a gene containing a sequence variation, according to the invention, can be detected by testing any tissue of a human subject. Human samples that are useful according to the invention include tissue or fluid samples containing a polynucleotide or polypeptide of interest, include but are not limited to plasma, serum, spinal fluid, lymph fluid, urine, stool, external secretions of the skin, respiratory, intestinal and genitoruinary tracts, saliva, blood cells, tumors, organs, tissue and samples of *in vitro* cell culture constituents. Genomic DNA, cDNA or RNA can be prepared from the human sample according to the methods described above.

P. Methods of Use

1. Nucleic Acid Diagnosis and Diagnostic Kits

In order to detect the presence of an allele of a gene predisposing an individual to

osteoarthritis, a biological sample such as blood is prepared and analyzed for the presence or absence of susceptibility alleles of a gene containing a polymorphism, according to the invention. Results of these tests and interpretive information will be returned to the health care provider for communication to the tested individual. Such diagnoses may be performed by diagnostic laboratories, or, alternatively, diagnostic kits are manufactured and sold to health care providers or to private individuals for self-diagnosis.

Initially, the screening method will involve amplification of the relevant gene sequences. In another preferred embodiment of the invention, the screening method involves a non-PCR based strategy. Such non-PCR based screening methods include Southern blot analysis to detect the presence of a variant form of a gene in a sample comprising total genomic DNA from the individual being tested. Alternatively, northern blot analysis can be used to detect an aberrant mRNA encoded by a gene, that exhibits altered stability or is the result of alternative splicing in a sample comprising RNA from an individual being tested. The methods of S1 nuclease analysis, RNase protection and primer extension can also be used to determine both the endpoint and the amount of a gene specific mRNA (Ausubel et al., supra). Both PCR and non-PCR based screening strategies can detect target sequences with a high level of sensitivity.

The preferred method, according to the invention, is target amplification. According to this method, the target nucleic acid sequence is amplified with polymerases. One particularly preferred method using polymerase-driven amplification is PCR (described above). The polymerase chain reaction and other polymerase-driven amplification assays can achieve over a million-fold increase in copy number through the use of polymerase-driven amplification cycles. PCR primers useful for target amplification according to the invention, will be designed to amplify a region of DNA containing one or more polymorphisms. Allele specific primers (comprising one or more polymorphisms) are also useful for detecting gene sequence variations by PCR methodologies according to the invention. The absence of a particular polymorphism will be indicated by the absence of an amplified product when the amplification step is carried out in the presence of allele specific primers. Once amplified, the resulting nucleic acid can be sequenced and the specific sequence of the test DNA will be compared with the wild type sequence by using the computer programs described in Section F entitled "Identification and Characterization of Polymorphisms". Alternatively, the amplified product will be analyzed by Southern blot assay with nucleic acid probes. Nucleic acid probes, useful according to the invention, will be specifically hybridizable to a mutant form of a gene but not to the wild type gene due to the presence of one or more polymorphisms.

When a probe comprising the target sequence, according to the invention, is used to detect the

presence of the target sequences via non PCR-based strategies, (for example, in screening for osteoarthritis susceptibility), the biological sample to be analyzed, such as blood or serum, may be treated, if desired, to extract the nucleic acids (as described above). The sample nucleic acids (isolated from a biological sample or amplified by PCR) may be prepared in various ways to facilitate detection of the target sequence; e.g. denaturation, restriction digestion, electrophoresis or dot blotting. Preferably, the targeted region of the nucleic acids being analyzed are at least partially single-stranded to form hybrids with the targeting sequence of the probe. If the sequence is naturally single-stranded, denaturation will not be required. However, if the sequence is double-stranded, the sequence will probably need to be denatured. Denaturation can be carried out by various techniques known in the art.

To detect the presence of a sequence variation in a gene, according to the invention, analyte nucleic acid and probe will be incubated under conditions which promote stable hybrid formation of the target sequence in the probe with the putative targeted sequence in the sample DNA. If the region of the probe which is used to bind to the analyte is designed to be completely complementary to the targeted region, high stringency conditions are desirable in order to prevent false positives. However, conditions of high stringency will be used only if the probes are complementary to regions of the chromosome which are unique in the genome. The stringency of hybridization is determined by a number of factors (described above). Detection, if any, of the resulting hybrid is usually accomplished by the use of labeled probes. Alternatively, the probe may be unlabeled, but may be detectable by specific binding with a ligand which is labeled, either directly or indirectly. Suitable labels, and methods for labeling probes and ligand are known in the art, and are described in Section C entitled "Production of a Nucleic Acid Probe".

Accordingly, the foregoing screening method may be modified to identify individuals having a gene containing a neutral polymorphism not associated with osteoarthritis, by preferably amplifying DNA fragments of a gene derived from a particular individual. The amplified DNA fragments are sequenced and the sequence is compared to the consensus gene sequence containing neutral polymorphisms. At this time, differences between the individual's coding sequence for a gene and a consensus sequence for the same gene are determined wherein the presence of any neutral polymorphisms and the absence of a polymorphisms not previously identified as neutral polymorphisms can be correlated with an absence of increased genetic susceptibility to osteoarthritis resulting from a mutation in a gene coding sequence.

In another embodiment of the invention, detection of a polymorphism will be performed by detecting loss of a restriction enzyme recognition site due to the presence of one or more

polymorphisms. According to this embodiment, a polymorphism will be detected with a polynucleotide probe that is capable of detecting a restriction enzyme fragment containing the polymorphism, wherein the fragment is of a size that can be easily separated on an agarose gel and visualized by Southern blot analysis. A polynucleotide probe according to this embodiment of the invention can be specific for a sequence within the candidate gene or outside of the candidate gene.

It is also contemplated within the scope of this invention that the nucleic acid probe assays of this invention will employ a mixture of nucleic acid probes capable of detecting a gene. Thus, in one example to detect the presence of a gene in a test sample, more than one probe complementary to a gene is employed and in particular the number of different probes is alternatively 2, 3, or 5 different nucleic acid probe sequences. In another example, to detect the presence of mutations in the gene sequence in a patient, more than one probe complementary to a gene is employed wherein the probe mixture includes probes capable of binding to the allele- specific mutations identified in populations of patients with alterations in a gene. In this embodiment, any number of probes can be used, and will preferably include probes corresponding to the major gene mutations identified as predisposing an individual to osteoarthritis.

Northern blot analysis, S1 nuclease analysis, RNase protection and primer extension (Ausubel et al., supra) are also methods according to the invention for detecting changes in mRNA resulting from the presence of one or more polymorphisms in the sequence of a gene.

Additionally, of the methods of genotyping described in Section F entitled "Identification and Characterization of Polymorphisms" can be used for diagnostics according to the invention.

2. Peptide Diagnosis and Diagnostic Kits

Osteoarthritis can also be detected on the basis of an alteration of the wild-type polypeptide. Such alterations can be determined by sequence analysis in accordance with conventional techniques. More preferably, antibodies (polyclonal or monoclonal) are used to detect differences in, or the absence of peptides derived from a gene of interest. The antibodies may be prepared as described above in Section I entitled "Preparation of Antibodies". Preferably, antibodies will immunoprecipitate the protein product of a gene from solution as well as react with the protein product of a gene on Western or immunoblots of polyacrylamide gels. Antibodies useful according to the invention will also detect the protein product of a gene in paraffin or frozen tissue sections, using immunocytochemical techniques.

Preferred embodiments relating to methods for detecting wild type or mutant forms of the protein product of a gene include enzyme linked immunosorbent assays (ELISA), radioimmunoassay

(RIA), immunoradiometric assays (IRMA) and immunoenzymatic assays (IEMA), including sandwich assays using monoclonal and/or polyclonal antibodies. Exemplary sandwich assays are described by David et al. In U.S. Pat. Nos. 4,376,110 and 4,486,530, hereby incorporated by reference.

3. Drug Screening

This invention is particularly useful for screening therapeutic compounds by using the mutant gene or protein product or binding fragment of the gene in any of a variety of drug screening techniques.

The protein product or fragment of a gene employed in such a test may either be free in solution, affixed to a solid support, expressed on the surface of a cell, or located intracellularly. One method of drug screening utilizes eukaryotic or procaryotic host cells which are stably transformed with a recombinant polynucleotide expressing the polypeptide or fragment, preferably in competitive binding assays. Such cells, either in viable or fixed form, can be used for standard binding assays. In particular, these cells can be used to measure formation of a complex comprising the protein product or fragment of a gene and the agent being tested. Alternatively, these cells can be used to determine if the formation of a complex between the protein product or fragment of a gene and a known ligand is interfered with by an agent being tested.

Thus, the present invention discloses methods useful for drug screening wherein such methods comprise contacting a candidate drug with a polypeptide or fragment derived from a gene and assaying (i) for the presence of a complex between the drug and the polypeptide derived or fragment derived from a gene, or (ii) for the presence of a complex between the polypeptide or fragment derived from a gene and a ligand, by methods well known in the art. Preferably, the polypeptide or fragment derived from a gene is labeled for use in competitive binding assays. Methods for producing a labeled protein by *in vitro* translation are described in Section J entitled "Preparation of a Labeled Protein". Free polypeptide or fragment will be separated from that present in a protein:protein complex, and the amount of free (i.e., uncomplexed) label will be used as a measure of the binding of the test drug to the polypeptide or the ability of the test drug to interfere with protein:ligand binding.

Another method of drug screening allows for high throughput screening for compounds exhibiting suitable binding affinity to the polypeptides and is described in detail in Geysen, WO 84/03564. According to this method, large numbers of different small peptide test compounds are synthesized on a solid substrate, such as plastic pins or another suitable surface. The peptide test compounds are reacted with the polypeptides or peptide fragments derived from a gene, and washed. Bound polypeptide is then detected by methods well known in the art.

Purified protein can be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies to the polypeptide can be used to capture the polypeptide or peptide fragment of interest and immobilize it on the solid support.

Competitive drug screening assays in which neutralizing antibodies capable of specifically binding the polypeptide of interest compete with a test compound for binding to the polypeptide or fragments thereof of interest are also useful according to the invention. According to this method, antibodies can be used to detect the presence of any test peptide which shares one or more antigenic determinants with the polypeptide of interest.

An additional technique for drug screening involves the use of host eukaryotic cell lines or cells (such as described above) which have a gene that produces a defective protein. According to this method, the host cell lines or cells are grown in the presence of a test drug compound. The rate of growth of the host cells is measured to determine if the compound is capable of regulating the growth of cells expressing a nonfunctional protein product of a gene. Alternatively, the ability of the test compound to restore the function of the mutant gene protein can be measured by using an appropriate *in vitro* assay for function of the protein product of a gene. Suitable *in vitro* functional assays are described in Section F entitled "Identification and Characterization of Polymorphisms". If the host cell lines or cells express a protein product of a gene that exhibits an aberrant pattern of cellular localization, the ability of the test compound to alter the cellular localization of the protein will be determined. Changes in the cellular localization of a protein of interest will be detected by performing cellular fractionation studies with biosynthetically labeled cells. Alternatively, the cellular localization of a protein of interest can be determined by immunocytochemical methods well known in the art.

A method of drug screening may involve the use of host eukaryotic cell lines or cells (described above) which have an altered gene that demonstrates an aberrant pattern of expression. By aberrant pattern of expression is meant the level of expression is either abnormally high or low, or the temporal pattern of expression is different from that of the wild type gene. The ability of a test drug to alter the expression of a mutant form of a gene can be measured by Northern blot analysis, S1 nuclease analysis, primer extension or RNase protection assays. Alternatively, if a mutant form of a gene contains an polymorphisms in the promoter region of a gene, cells can be engineered to express a reporter construct comprising a mutant gene promoter driving expression of a reporter gene (e.g. CAT, luciferase, green fluorescent protein). These cells can be grown in the presence of a test compound and the ability of a test compound to alter the level of activity of the mutant gene promoter can be determined by standard assays for each reporter gene which are well known in the art.

Candidate Drugs

A "candidate drug" as used herein, is any compound with a potential to modulate a phenotype associated with a particular disease according to the invention.

A candidate drug is tested in a concentration range that depends upon the molecular weight of the drug and the type of assay. For example, for inhibition of protein/protein complex formation, small molecules (as defined below) may be tested in a concentration range of 1 pg - 100 mg/ml, preferably at about 100 pg - 10 ng/ml; large molecules, e.g., peptides, may be tested in the range of 10 ng - 100 mg/ml, preferably 100 ng - 10 mg/ml.

Candidate drug compounds from large libraries of synthetic or natural compounds can be screened. Numerous means are currently used for random and directed synthesis of saccharide, peptide, and nucleic acid based compounds. Synthetic compound libraries are commercially available from a number of companies including Maybridge Chemical Co. (Trevillet, Cornwall, UK), Comgenex (Princeton, NJ), Brandon Associates (Merrimack, NH), and Microsource (New Milford, CT). A rare chemical library is available from Aldrich (Milwaukee, WI). Combinatorial libraries are available and can be prepared. Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts are available from e.g., Pan Laboratories (Bothell, WA) or MycoSearch (NC), or are readily produceable by methods well known in the art. Additionally, natural and synthetically produced libraries and compounds are readily modified through conventional chemical, physical, and biochemical means.

Useful compounds may be found within numerous chemical classes, though typically they are organic compounds, and preferably small organic compounds. Small organic compounds have a molecular weight of more than 50 yet less than about 2,500 daltons, preferably less than about 750 daltons, more preferably less than about 350 daltons. Exemplary classes include heterocycles, peptides, saccharides, steroids, and the like. The compounds may be modified to enhance efficacy, stability, pharmaceutical compatibility, and the like. Structural identification of an agent may be used to identify, generate, or screen additional agents. For example, where peptide agents are identified, they may be modified in a variety of ways to enhance their stability, such as using an unnatural amino acid, such as a D-amino acid, particularly D-alanine, by functionalizing the amino or carboxylic terminus, e.g. for the amino group, acylation or alkylation, and for the carboxyl group, esterification or amidification, or the like.

Determination of Activity of a Drug

A candidate drug, assayed according to the invention as described above, is determined to be

effective if its use results in a change of about 10% of a phenotype associated with a disease according to the invention.

The level of modulation by a candidate modulator of a phenotype associated with a disease according to the invention, may be quantified using any acceptable limits, for example, via the following formula, which describes detections performed with a radioactively labeled probe (e.g., a radiolabeled antibody in an immunobinding experiment or a radiolabeled nucleic acid probe in a Northern hybridization).

$$\text{Percent Modulation} = \frac{(CPM_{Control} - CPM_{Sample})}{(CPM_{Control})} \times 100$$

where $CPM_{Control}$ is the average of the cpm in antibody/ligand complexes or on Northern blots resulting from assays that lack the candidate modulator (in other words, untreated controls), and CPM_{Sample} is the cpm in antibody/ligand complexes or on Northern blots resulting from assays containing the candidate modulator. A similar calculation is performed where the assay comprises use of a labeling system or system of measuring enzymatic activity in which there is a linear relationship between the amount of label detected and the amount of protein or nucleic acid being represented per unit of label or the amount of protein or nucleic acid represented by a unit of enzymatic activity.

4. Rational Drug Design

Rational drug design is useful for producing either structural analogs of biologically active polypeptides of interest or small molecules with which polypeptides of interest interact (e.g., agonists, antagonists, inhibitors) in order to design drugs which are, for example, more active or stable forms of the polypeptide, or which enhance or interfere with the function of a polypeptide *in vivo*. See, e.g., Hodgson, 1991, BioTechnology, 9:19. According to one method of rational drug design, the three-dimensional structure of a protein of interest (e.g., the polypeptide product of the gene) or, or the complex comprising the protein product of a gene in association with its ligand, is determined by x-ray crystallography, by computer modeling or most typically, by a combination of approaches. Alternatively, useful information regarding the structure of a polypeptide may be obtained by modeling based on the structure of homologous proteins. Rational drug design has been used successfully in the development of HIV protease inhibitors (Erickson et al., 1990, Science, 249: 527).

Rational drug design may also involve the analysis of peptides derived from the protein

product of a gene by an alanine scan (Wells, 1991, Methods in Enzymol., 202: 390). According to this method, each of the amino acid residues of the peptide is sequentially replaced by alanine, and the effect of this amino acid substitution on the peptide's activity is determined. This technique can be used to determine the functionally relevant regions of the peptide.

Another experimental approach to rational drug design will involve the isolation of a target-specific antibody (selected by a functional assay) and the determination of the crystal structure of this antibody. Theoretically, this approach will yield a pharmacore upon which subsequent drug design can be based. Alternatively, if anti-idiotypic antibodies (anti-ids) specific for a functional, pharmacologically active antibody are generated, there is no need to determine the crystallographic structure of the target-specific antibody. It is expected that the binding site of the anti-ids will be an analog of the original receptor. The anti-id could then be used to identify and isolate potentially therapeutic peptides from banks of chemically or biologically produced banks of peptides. These selected peptides would then function as pharmacores.

According to these methods it may be possible to design drugs which demonstrate increased activity or stability of the protein product of a gene or which function as inhibitors, agonists, antagonists, etc. of the activity of a protein product of a gene. The availability of cloned gene sequences, including polymorphisms, ensures that sufficient amounts of the polypeptide product of a gene are available to facilitate analytical studies such as x-ray crystallography. Furthermore, the knowledge of the sequence of the protein product of a gene provided herein will guide those using computer modeling techniques in place of, or in addition to x-ray crystallography.

5. Gene Therapy

The present invention also provides a method of supplying wild-type gene function to a cell which carries a mutant allele of a gene. By replacing a mutant gene with a wild type gene, it may be possible to reverse the symptoms of osteoarthritis in the recipient cells. a full length version of the wild-type gene, or a fragment of the gene, may be introduced into the cell in a vector such that the gene remains extrachromosomal and is expressed by the cell from the extrachromosomal location. More preferably, following introduction into the mutant cell, the wild-type gene or gene fragment should recombine with the endogenous mutant gene X already present in the cell. Such recombination requires a double recombination event which results in the correction of the gene mutation. Vectors for introduction of genes both for recombination and for extrachromosomal maintenance are known in the art, and any suitable vector may be used. Methods for introducing DNA into cells such as electroporation, calcium phosphate coprecipitation and lipofection are known in the art (described

above). Cells transformed with the wild-type gene can be used as model systems to study changes in the intensity of symptoms associated with osteoarthritis and drug treatments which promote such changes.

As generally discussed above, a gene or a fragment thereof, where applicable, may be used in gene therapy methods in order to increase the amount of the expression products of such genes in cells of patients with osteoarthritis. It may also be useful to increase the level of expression of a gene even in those cells in which the mutant gene is expressed at a "normal" level, but the gene product is not fully functional.

It other embodiments of the invention it may be useful to increase the amount of the expression products of a mutant form of a gene in a cell that expresses the wild type protein. Gene therapy can be carried out according to generally accepted methods, for example, as described by Friedman, 1991, In Therapy for Genetic Diseases; T. Friedman ed., Oxford University Press, pp. 105-121). Initially, the appropriate cells from a patient with osteoarthritis would be analyzed by the diagnostic methods described above, to determine the level of production of a polypeptide from a gene and the activity of a polypeptide product of a gene. A virus or plasmid vector (see further details below), comprising a copy of a gene and suitable expression control elements, and capable of replicating inside the cells, will be prepared. Suitable vectors are known and are disclosed in U.S. Pat. No. 5,252,479 and PCT published application WO 93/07282. The vector will be injected into the patient, either locally at an appropriate site according to the invention or systemically.

Gene transfer systems known in the art may be useful in the practice of the gene therapy methods of the present invention. These include viral and nonviral transfer methods. a number of viruses have been used as gene transfer vectors, including papovaviruses, e.g., 5V40 (Madzak et al., 1992, J Gen Virol., 73:1533), adenovirus (Berkner, 1992, Curr. Top. Microbiol. Immunol., 158:39; Berkner et al., 1988, BioTechniques, 6:616; Gorziglia and Kapikian, 1992, J Virol., 66:4407; Quantin et al., 1992, Proc. Natl. Acad. Sci. USA, 89:2581; Rosenfeld et al., 1992, Cell, 68:143 ; Wilkinson et al., 1992, Nucleic Acids Res. 20:2233; Stratford-Perricaudet et al., 1990, Hum. Gene Ther., 1:241), vaccinia virus (Moss, 1992, Curr. Top. Microbiol. Immunol., 158:25) adeno-associated virus (Muzyczka, 1992, Curr. Top. Microbiol. Immunol., 158:97; Ohi et al., 1990, Gene, 89:279), herpesviruses including HSV and EBV (Margolskee, 1992, Curr. Top. Microbiol. Immunol., 158:67, Johnson et al., 1992, J. Virol., 66:2952; Fink et al., 1992, Hum. Gene Ther., 3:11; Breakfield and Geller, 1987, Mol. Neurobiol., 1:337; Freese et al., 1990, Biochem. Pharmacol., 40: 2189), and retroviruses of avian (Brandyopadhyay and Temin, 1984, Mol. Cell. Biol., 4:749; Petropoulos et al., 1992, J. Virol., 66:3391), marine (Miller, 1992, Curr. Top. Microbiol. Immunol., 158:1; Miller et al., 1985, Mol. Cell.

Biol., 5:431; Sorge et al., 1984, Mol. Cell. Biol., 4:1730; Mann and Baltimore, 1985, J. Virol., 54:401; Miller et al., 1988, J. Virol., 62:4337), and human origin (Shimada et al., 1991, J. Clin. Invest., 88:1043); Helseith et al., 1990, J. Virol., 64:24 16; Page et al., 1990, J. Virol., 64: 5370; Buchschacher and Panganiban, 1992, J. Virol., 66:2731). Most human gene therapy protocols have been based on disabled murine retroviruses.

Nonviral gene transfer methods known in the art include chemical techniques such as calcium phosphate coprecipitation (Graham and van der Eb, 1973, Virology, 52:456; Pellicer et al., 1980, Science, 209:1414); mechanical techniques, for example microinjection (Anderson et al., 1980, Proc. Natl. Acad. Sci. USA, 77: 5399; Gordon et al., 1980, Proc. Natl. Acad. Sci.. USA, 77: 7380; Brinster et al., 1981, Cell, 27:223; Constantini and Lacy, 1981, Nature, 294:92); membrane fusion-mediated transfer via liposomes (Felgner et al., 1987, Proc. Natl. Acad. Sci. USA, 84:7413; Wang and Huang, 1989, Biochemistry, 28:9508; Kaneda et al. 1989, J. Biol. Chem., 264:12126; Stewart et al., 1992, Hum. Gen. Ther., 3:267; Nabel et al., 1990, Science, 249:1285; Lim et al., 1992, Circulation, 83:2007); and direct DNA uptake and receptor-mediated DNA transfer (Wolff et al., 1990, Science, 247:1465; Wu et al., 1991, J. Biol. Chem., 266:14338; Zenke et al., 1990, Proc. Natl. Acad. Sci. USA, 87:3655; Wu et al., 1989b, J. Biol. Chem., 264:16985; Wolff et al., 1991, BioTechniques, 11:474; Wagner et al., 1990, Proc. Natl. Acad. ScLUSA, 87:3410; Wagner et al., 1991, Proc. Natl. Acad. Sci.USA, 88:4255; Cotten et al., 1990, Proc. Natl. Acad. Sci.USA, 87:4033; Curiel et al., 1991a, Proc. Natl. Acad. Sci.USA, 88:8850; Curiel et al., 1991b, Hum. Gene Ther., 3:147.

In an approach which combines biological and physical gene transfer methods, plasmid DNA of any size is combined with a polylysine-conjugated antibody specific to the adenovirus hexon protein, and the resulting complex is bound to an adenovirus vector. The trimolecular complex is then used to infect cells. The adenovirus vector permits efficient binding, internalization, and degradation of the endosome before the coupled DNA is damaged.

Liposome/DNA complexes have been shown to be capable of mediating direct *in vivo* gene transfer. While in standard liposome preparations the gene transfer process is nonspecific, localized *in vivo* uptake and expression have been reported in tumor deposits, for example, following direct *in situ* administration (Nabel, 1992, Hum. Gen. Ther., 3:399).

Gene transfer techniques which target DNA directly to an appropriate tissue, e.g., a tissue that normally expresses the protein product of the candidate gene of the invention, is preferred. Receptor-mediated gene transfer, for example, is accomplished by the conjugation of DNA (usually in the form of covalently closed supercoiled plasmid) to a protein ligand via polylysine. Ligands are chosen on the basis of the presence of the corresponding ligand receptors on the cell surface of the

target cell/tissue type. These ligand-DNA conjugates can be injected directly into the blood if desired and are directed to the target tissue where receptor binding and internalization of the DNA-protein complex occurs. To overcome the problem of intracellular destruction of DNA, coinfection with adenovirus can be included to disrupt endosome function.

6. Peptide Therapy

Peptides which have gene activity can be supplied to cells which carry mutant or missing alleles of a gene. Alternatively, peptides specific for a mutant form of the protein product of a gene can be supplied to cells carrying a wild type protein. The protein product of a gene can be produced by expression of the cDNA sequence in bacteria, for example, using known expression vectors (as described in Section H entitled "Production of a Mutant Protein"). Alternatively, the protein product of a gene can be extracted from mammalian cells engineered to produce the protein product of a gene of interest. In addition, the techniques of synthetic chemistry can be employed to synthesize the protein product of a gene. Any of the above techniques can provide a preparation of protein product of a gene that is substantially free of other human proteins. This is most readily accomplished by carrying out protein synthesis in a microorganism or *in vitro*.

Active gene molecules can be introduced into cells by microinjection or by the use of liposomes, for example. Alternatively, some active molecules may be taken up by cells, actively or by diffusion. Extracellular application of the protein product of a gene may be sufficient to decrease or reverse the physiological effects of osteoarthritis. Other molecules with the activity of a protein product of a gene (for example, peptides, drugs or organic compounds) may also be used to effect such a reversal. Modified polypeptides having substantially similar function may also be useful for peptide therapy.

7. Transformed Hosts

Cells and animals which carry a mutant allele of a gene can be used as model systems to study and test for substances which have potential as therapeutic agents. Following application of a test substance to the cells, the phenotype of the cell will be determined. Any variety of phenotypic changes associated with osteoarthritis can be assessed, including insulin resistance and combined insulin resistance/insulin secretion detect. Assays for each of these traits are known in the art.

Animals useful for testing therapeutic agents can be selected after mutagenesis of whole animals or after treatment of germline cells or zygotes. Such treatments include insertion of mutant alleles of a gene, usually from a second animal species, as well as insertion of disrupted homologous

genes. Alternatively, the endogenous gene of the animals may be disrupted by insertion or deletion mutation or other genetic alterations using conventional techniques (Capecchi, 1989, Science, 244:1288; Valancius and Smithies, 1991, Mol. Cell. Biol., 11:1402; Hasty et al., 1991, Nature, 350:243; Shinkai et al., 1992, Cell, 68:855; Mombaerts et al., 1992, Cell, 68:869; Philpott et al., 1992, Science, 256:1448; Snouwaert et al., 1992, Science, 257:1083; Donehower et al., 1992, Nature, 356;215). Following the administration of test substances, the physiological changes associated with osteoarthritis will be assessed. If the test substance prevents or suppresses any of these physiological changes, then the test substance will be considered a candidate therapeutic agent for the treatment of osteoarthritis. These animal models provide an extremely important testing vehicle for potential therapeutic products.

8. Use of a Polynucleotide as a Unique Sequence Marker:

Polynucleotides can be used to mark objects or substances for the purposes of later identification. Thus, polynucleotides of the invention are useful for tracking the manufacture and distribution of a large number of diverse substances, including but not limited to: (1) natural resources such as animals, plants, oil, minerals, and water; (2) chemicals such as drugs, solvents, petroleum products, and explosives; (3) commercial by-products including pollutants such as radioactive or other hazardous waste; and (4) articles of manufacture such as guns, typewriters, automobiles and automobile parts. A nucleic acid according to the invention, when used as a marker, thus aids in the determination of product identity and so provides information useful to manufacturers and consumers.

Polynucleotides have the advantage over other marking materials of being readily amplifiable through the use of polymerase chain reaction (PCR) technology. The method of PCR is well known in the art. PCR is performed as described by Mullis & Faloona, 1987, Methods Enzymol., 155:335, herein incorporated by reference. It is the unique sequence of a polynucleotide which renders it useful as a marker, since thesequence, or a characteristic pattern derived from its sequence, confers a property on the polynucleotide which permits it to be tracked.

It is contemplated that a novel polynucleotide sequence of the invention, or fragments or derivatives of it may be used as markers by their attachment to or mixture in objects or substances to be marked. Methods for marking various classes of substances and later detection of the tags in those substances are disclosed in U.S. Patent Nos. 5,451,505, and 5,643,728.

Briefly, the use of a polynucleotide of the invention as a marker may entail combining a polynucleotide with the substance or object to be marked, using methods appropriate to that substance or object; and detecting the marker through amplification of the polynucleotide sequence using PCR technology, followed by either sequence analysis or identification by other means known in the art

(e.g., hybridization assays).

The methods of applying a marker nucleic acid to a substance or object and subsequent detection of that nucleic acid will vary depending upon the nature of the substance or object and the environment to which it will be exposed. For example, inert solids such as paper, many pharmaceutical products, wood, some foodstuffs, etc., can be either processed with the marker nucleic acid, or the nucleic acid may be sprayed onto their surfaces. Chemically active substances, such as foodstuffs with enzymatic activity, polymers with charged groups, or acidic pharmaceuticals may require that a protective composition (e.g., liposomes) be added to the nucleic acid being used as a marker.

In order to mark liquids, the nucleic acid may be mixed directly with the liquid, or, if the chemical nature of the liquid is not compatible with this approach (i.e., nucleic acids are not soluble in the liquid), the nucleic acid may be mixed with a detergent to enhance its solubility. Containerized gases may be marked simply by adding a nucleic acid to the container in dry form, as it will be dispersed throughout the gas as the gas is released.

The amount of nucleic acid to add to a substance as a marker will also vary with the given situation, as will the detection strategy. PCR technology, however, allows the amplification and detection of as little as one molecule from a sample. Other means of detection, such as hybridization assays require that more nucleic acid be recovered from a sample to efficiently detect it. PCR can be combined with a hybridization assay, however, to enhance the sensitivity of the method.

A nucleic acid sequence used as a marker will generally be from 20 to 1,000 bases long, and preferably will be 60 to 1,000 bases long when PCR is to be used to detect the marker.

One example of a substance for which nucleic acid marking is suited is gunpowder. Marked gunpowder may be prepared as follows: 1) add 16 ng of nucleic acid bearing the chosen marker sequence (derived from a polynucleotide of the invention) to 1 ml of distilled water; 2) mix the solution of nucleic acid with 1 g of nitrocellulose-based gunpowder; and 3) dry in air or under vacuum at 85^{0}C. To recover the marker from gunpowder: 1) wash the gunpowder sample with 1 ml of distilled water; 2) add 50 m1 of the wash solution to a standard PCR mix, or, alternatively, place gunpowder flakes directly into a 100 ml PCR mix; and 3) amplify according to standard PCR methods using primers which anneal at opposite ends and on opposite strands of the sequence used as a marker (annealing and extension conditions will depend upon the exact sequences chosen for oligonucletide primers, and may be adjusted according to methods known in the art).

Another example of a substance which may be marked with a nucleic acid according to the invention is ink. To prepare marked ink sample: 1) if the ink is water insoluble, mix the nucleic acid with detergents as for oil. If the ink is water soluble, add nucleic acid directly to the ink to a

concentration of about 1 to 20 ng per ml. To recover the marker from ink, proceed as for oils and medicines.

In the above examples, the presence of an amplification product of the proper size (visualized, for example by gel electrophoresis alongside nucleic acid size markers followed by ethidium bromide staining of the gel, according to standard methods) will indicate the presence of the marker in the sample. In some instances, the PCR product may be further subjected to hybridization analysis or to sequencing to enhance the accuracy of the method. A method of hybridization analysis which can be used is described herein.

9. Use of a Polynucleotide of the Invention as a Marker for Chromosome Mapping:

Because a polynucleotide of the invention is novel, (that is, its sequence is unique),it is useful as a marker for chromosomal mapping. There are a number of methods of chromosomal mapping known in the art. Prominent among them is the variant of the *in situ* hybridization technique known as "Fluorescence *In Situ* Hybridization", or FISH. Details of methods and solutions used for *in situ* hybridization are well-known in the art. There are many variations of the FISH technique itself, however the basic approach is similar in each case. Essentially, *in situ* hybridization of cells, nuclei, or metaphase chromosome spreads is performed with a polynucleotide probe either directly labeled with a fluorochrome, or labeled with a moiety which will be bound by a fluorochrome tagged entity. The hybridized probe is visualized by irradiation of the sample with light in the wavelength which excites fluorescence from the fluorochrome. When combined with standard methods of karyotyping known in the art, this method allows the polynucleotide sequence to be localized to a particular arm of a particular chromosome. Once mapped to a specific chromosome, the location of the novel polynucleotide sequence on that chromosome may be further localized by *in situ* hybridization along with probes specific for known genes or sequences, labeled with other fluorescent tags which allow the differentiation of the signals from the different probes. Such an approach and various adaptations of it allows the localization of the novel gene relative to a known gene. Methods of generating and using fluorescence-labeled polynucleotide probes for FISH and chromosome mapping are known in the art (for example, see Malcolm et al., 1981, Ann. Hum. Genet., 45:134; Bar-Am et al., 1992, Genes. Chromosomes & Cancer, 4:314; Pinkel et al., 1988, Proc. Natl. Acad. Sci. USA, 85:9138; U.S. Patent No. 5,728,527). Additional variations of the chromosome mapping method utilize a PCR approach (Dionne et aL, 1990, BioTechniques, 8(2):190 and Iggo et al., 1989, Proc. Natl. Acad. Sci. USA, 86:6211).

In addition to being able to determine the chromosomal location of the novel polynucleotide,

similar technology, in which FISH is combined with flow cytometry, will allow the polynucleotide of the invention to be used to sort chromosomes, nuclei, or whole cells containing various dosages (i.e., gene copy numbers) of the gene encoding that polynucleotide (Hulfdin et al., 1998, Nuc. Acids Res., 26:3651). The novel polypeptide may also be useful as a diagnostic indicator of a disease, including but not limited to those listed in Table I (Kuo et al., 1990, Am. J. Hum. Genet., 47:A119).

10. Use of a Polynucleotide of the Invention as a Marker for Analysis of Forensic Materials

Forensic science depends heavily on methods for determining the source of various compounds associated with criminal activity. In particular, the identification of individuals involved in criminal activity through analysis of substances found at the crime scenes is critical. Such identification is possible with genetic typing, which involves the determination of the genotype of an individual with regard to loci which are polymorphic within the population. As used herein, "polymorphic" refers to a gene or other segment of DNA which shows nucleotide sequence variability from individual to individual. The use of PCR techniques and nucleotide probes to detect even single nucleotide changes in a polynucleotide sequence has revolutionized the field of forensic serology (see Reynolds and Sensabaugh, 1991, Anal. Chem., 63:2). For an example of polymorphisms useful for forensic identification and methods of typing samples with regard to those polymorphisms, see U.S. Patent # 5,273,883.

If a polynucleotide of the invention is found to have nucleotide sequence variation among individuals within a population, it may be useful in the analysis of forensic samples. There are a number of methods known to those skilled in the art for typing nucleic acids with regard to polymorphisms. It should be understood that any such method is acceptable according to the invention. One particular method is termed the "reverse dot blot" method. The basic steps involved are: 1) oligonucleotides bearing the sequences of various polymorphic forms of the polynucleotide region to be analyzed are bound to membranes; 2) labeled, PCR-amplified fragments, derived from the sample to be genotyped, and corresponding to the polymorphic region ("target DNA") are allowed to hybridize to the bound oligonucleotides under conditions which only allow the hybridization of molecules with 100% complementary sequences; 3) unbound target DNA is removed; and 4) hybridized molecules are detected.

The specific genotype of the individual from whom the target sample was obtained (amplified), with regard to the polymorphic region of a polynucleotide of the invention, may thus be determined by screening a panel of probes containing the known polymorphic sequence variations of

that region. It should be understood that the hybridization conditions may be adjusted by one of skill in the art so that limited amounts of non-complementarity, including single base mismatches, may be detected with this method.

Q. Pharmaceutical Compositions--Prevention and Treatment

1. Administration of Pharmaceutical Compositions

Administration of pharmaceutical compositions is accomplished orally or parenterally. Methods of parenteral delivery include topical, intra-arterial (directly to the tumor), intramuscular, subcutaneous, intramedullary, intrathecal, intraventricular, intravenous, intraperitoneal, or intranasal administration. In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically acceptable carrier preparations which can be used pharmaceutically.

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combination of active compounds with solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are carbohydrate or protein fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethyl cellulose; and gums including arabic and tragacanth; and proteins such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate.

Dragee cores are provided with suitable coatings such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, ie, dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders such as lactose or

starches, lubricants such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations for parenteral administration include aqueous solutions of active compounds. For injection, the pharmaceutical compositions of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hank's solution, Ringer' solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated or used in the formulation. Such penetrants are generally known in the art.

2. Manufacture and Storage

The pharmaceutical compositions of the present invention may be manufactured in a manner that known in the art, e.g. by means of conventional mixing, dissolving, granulating, dragee-making, levitating, emulsifying, encapsulating, entrapping or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to hydrochloric, sulfuric, acetic, lactic, tartaric, malic, succinic, etc... Salts tend to be more soluble in aqueous or other protonic solvents that are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder in 1mM-50 mM histidine, 0.1%-2% sucrose, 2%-7% mannitol at a PhRange of 4.5 to 5.5 that is combined with buffer prior to use.

After pharmaceutical compositions comprising a compound of the invention formulated in a acceptable carrier have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition with information including amount, frequency and method of administration.

3. Therapeutically Effective Dose

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose.

The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, or in animal models, usually mice, rabbits, dogs, or pigs. The animal model is also used to achieve a desirable concentration range and route of administration. Such information can then be use to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of protein or its antibodies, antagonists, or inhibitors which ameliorate the symptoms or conditions. Therapeutic efficacy and toxicity of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, eg, ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population). The dose ratio between therapeutic and toxic effects is the therapeutic index, and it can be expressed as the ratio, LD50/ED50. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animals studies is used in formulating a range of dosage for human use. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED5O with little or no toxicity. The dosage varies within this range depending upon the dosage from employed, sensitivity of the patient, and the route of administration.

The exact dosage is chosen by the individual physician in view of the patient to be treated. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Additional factors which may be taken into account include the severity of the disease state; age, weight and gender of the patient; diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long acting pharmaceutical compositions might be administered every 3 to 4 days, every week, or once every two weeks depending on a half-life and clearance rate of the particular formulation.

Dosage amounts may vary from 0.1 to 100,000 micrograms per person per day, for example, lug, lOug, lOOug, 500 ug, 1mg, 10mg, and even up to a total dose of about 1g per person per day, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature. See U.S. Patent Nos. 4,657,760; 5,206,344; or 5,225,212, hereby incorporated by reference. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotide or polypeptides will be specific to particular cells, conditions, locations, etc...

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way

whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, including U.S. Ser. No. 60/342,603, are hereby expressly incorporated by reference.

EXAMPLES

1. Establishment of an Association Between a Given Polynucleotide Sequence and Diabetes

A polynucleotide sequence according to the invention containing a mutation which is believed to be associated with a disease, can be statistically linked to that disease by linkage analysis. An animal model system exhibiting a particular phenotypic defect that is characteristic of the disease of interest is selected. A series of genetic crosses is performed in this animal model system between individuals having an observable mutant phenotype and normal individuals of a control strain. At least one disease-related locus or a chromosomal marker that does not comprise a disease related locus is used as a marker in these crosses. If a statistically significant pattern of non-random assortment of the mutant trait with a marker locus is observed, the trait is linked to the marker locus.

Similarly, linkage analysis can be performed on an existing human or other mammalian pedigree. According to this method, numerous genetic loci from affected and unaffected family members are compared. Non-random assortment of a given genetic marker between affected and unaffected family members relative to the distributions observed for other genetic loci indicates that the marker (for example, a variant isoform of a gene) either contributes to the disease or is in physical proximity to another that does so.

If either approach demonstrates a non-random assortment of the disease-related phenotype with a marker locus, this is indicative of an association between the gene underlying the defect and that locus. Because the strength of any conclusion drawn from linkage analysis is statistically-based, the accuracy of the results is thought to be proportional to the number of crosses or family members and genetic loci analyzed.

2. Screening Assay For a Disease

A polynucleotide sequence according to the invention can be used as a marker for a normal phenotype or for a phenotype associated with a disease of interest.

If it can be demonstrated by the methods of phenotyping, described above, that a particular sequence is associated with a disease phenotype, this sequence can be used as a marker for a

particular disease. A sequence of interest can be used as a probe to screen genomic DNA from individuals by Southern blot analysis according to the method described above. If the sequence of interest is detected by Southern blot analysis, and the presence of this sequence is confirmed by direct sequencing, it can be concluded that the individual from which the genomic DNA has been isolated has an increased frequency for the development of the disease for which the sequence is a marker.

The marker can also be used as a disease indicator according to the method of PCR. A genomic DNA sample of interest can be analyzed in a PCR reaction wherein one of the primers contains the marker sequence. If the marker sequence is present in the sample DNA, a PCR product will be produced. Alternatively, the PCR primers can be designed such that they amplify a region containing the marker sequence. The amplified product can be analyzed by hybridization methods, described above, to determine the presence of the sequence of interest.

3. Use of a Given Polynucleotide as a Target for Drug Screening

A polynucleotide according to the invention, containing a mutation which is believed to be associated with a disease can be used a target for drug screening.

One method of drug screening utilizes eukaryotic or procaryotic host cells which are stably transformed with a polynucleotide according to the invention and either exhibit a particular phenotype characteristic of the presence of the polynucleotide or express a polypeptide or fragment encoded by the polynucleotide. Such cells, either in viable or fixed form, can be used for standard competitive binding assays. In particular, these cells can be used to measure formation of a complex comprising the protein product or fragment of a polynucleotide according to the invention and the agent being tested. Alternatively, these cells can be used to determine if the formation of a complex between the protein product or fragment of a polynucleotide according to the invention and a known ligand is interfered with by an agent being tested.

An alternative method for drug screening involves using of eukaryotic cell lines or cells (such as described above) which contain a polynucleotide according to the invention that produces a defective protein. According to this method, the host cell lines or cells are grown in the presence of a test drug. The rate of growth of the host cells is measured to determine if the compound is capable of regulating the growth of cells expressing a nonfunctional protein product of the polynucleotide according to the invention. Preferably, a drug that is useful according to the invention will increase or decrease the growth rate of a cell by at least 10%. Alternatively, the ability of the test compound to restore the function of the mutant gene protein by at least 10% can be measured by using an appropriate *in vitro* assay for function of the protein product of a gene (as described in Section F

entitled “Identification and Characterization of Polymorphisms”). If the host cell lines or cells express a protein product of a gene that exhibits an aberrant pattern of cellular localization, the ability of the test compound to alter the cellular localization of the protein by at least 10% will be determined. Changes in the cellular localization of a protein of interest will be detected by performing cellular fractionation studies with biosynthetically labeled cells. Alternatively, the cellular localization of a protein of interest can be determined by immunocytochemical methods well known in the art.

A method of drug screening may also involve the use of host eukaryotic cell lines or cells (described above) which have an altered gene that demonstrates an aberrant pattern of expression. By aberreant pattern of expression is meant the level of expression is either abnormally high or low, or the temporal pattern of expression is different from that of the wild type gene. The ability of a test drug to alter the expression of a mutant form of a gene by at least 10% can be measured by Northern blot analysis, S1 nuclease analysis, primer extension or Rnase protection assays, as described above. Alternatively, if a mutant form of a gene contains a polymorphism in the promoter region of a gene, cells can be engineered to express a reporter construct comprising a mutant gene promoter driving expression of a reporter gene (e.g. CAT, luciferase, green fluorescent protein). These cells can be grown in the presence of a test compound and the ability of a test compound to alter the level of activity of the mutant gene promoter can be determined by standard assays for each reporter gene which are well known in the art.

A transgenic animal whose genomic DNA contains a polynucleotide associated with a particular phenotypic defect that is characteristic of the disease of interest, and a normal, control anomal (not containing the polynucleotide) can be treated with a candidate drug according to the invention. The ability of a candidate drug to ameliorate symptoms of the disease, by at least 10%, will be analyzed by assessing the disease syptoms and their amelioration.

4. Selection of Osteoarthritis Candidate Gene Set

Genes involved in osteoarthritis

Key pathogenic processes involved in osteoarthritis are:

1. chondrocyte differentiation, development, apoptosis and signalling
2. cartilage components and synthesis : proteoglycans, hyaluronan synthases, extracellular matrix molecules

3. cartilage degradation: cathepsin proteases and matrix metalloproteinases, their inhibitors
4. bone remodelling signals (e.g. RANK/RANKL): BMPs, TGFbeta, interleukins, their receptors and antagonists, downstream signaling.
5. synovial fluid components
6. systemic factors influencing bone and cartilage remodelling: leptin, estrogen, progesterone, inflammatory cytokines, retinoic acid

Polymorphisms at the following genes have been reported in the literature to be involved with increased risk of osteoarthritis. They include components of the extracellular matrix, and bone-remodelling signalling components (Table 2)

With the aim of expanding and improving on the current limited knowledge of osteoarthritis genetic predisposition, we have collected over 500 candidate bone and cartilage remodelling genes using the following methods:

1. extensive literature search for genes involved in relevant biochemical pathways and physiological processes

2. analysis and comparisons of cDNA libraries within the Incyte Lifeseq® database from relevant normal and diseased tissues and *in vitro* modelling systems

3. co-expression analysis using Incyte's "Guilt by Association" algorithm which identifies novel genes in key biochemical pathways by comparing the expression patterns of genes within the Lifeseq® database

5. Polymorphisms in Genes Associated with Osteoarthritis

The osteoarthritis candidate gene list was compiled using gene or gene sequences selected from literature sources, using sequence homology, library subtraction and expression analysis.

Expression analysis was performed using "quilt-by-association" queries to identify Incyte-novel and known genes not previously associated with diabetes which have similiar expression patterns to genes known to be involved in diabetes or related conditions. Guilt-by-association analysis was performed as described in Walker et al. 1999 Genome Res 9:1198; Walker et al. 1999 Ismb :282; and US Patent Application 09/226,994 entitled "Insulin-Synthesis Genes" (Atty Docket No: PB-0008

US) filed January 7, 1999, all of which are incorporated by reference.

Polymorphism discovery was by fSSCP as decribed in section F "Identication and Characterization of Polymorphisms", subsection b5 for polymorphisms referred to in Table 3 for source wetSNPs. Polymorphisms referred to as source isSNPs were discovered as described in section F "Identification and Characterization of Polymorphisms", subsection a. Polymorphisms referred to as source dbSNPs are polymorphisms in public genomic sequence where gene structure is unknown. The polymorphisms were mapped to cDNA sequences in the LifeSeqGold database (Incyte) to identify gene identity.

6. Frequency of Polymorphisms in Diabetes Associated Genes and Polynucleotides in Various Populations

Polymorphisms identified in EXAMPLES 4 and 5 were genotyped against populations described below by fSSCP or FP-TDI as described above. The results of the population frquency studies are given in Table 2.

Two panels of human DNA have been developed to support the identification of frequent SNPs within an ethnically diverse population. The genomic Human Diversity Panel will be used where full genomic structure is available, and allows screening of the open reading frame of the gene, including splice junctions. In instances where genomic structure for selected candidate genes may not be available, a cDNA version of the HDP Screening Panel permits screening of the open reading frame of the gene.

This DNA panel is derived from 47 consented individuals from four ethnic groups (Caucasian, African-American, Asian and Hispanic). The panel is sufficiently sized to enable identification of 95% of SNPs with allele population frequencies >= 5%. Comparable utility of our panel with the NIH Diversity panel was demonstrated by parallel screening of 90 kilobases of coding sequence from each panel.

A cDNA counterpart to our Human Diversity Panel has been generated from lymphoblastoid cell lines to obviate the need for intron/exon structure in 50% of human genes. In the absence of genomic structure, this methodology will be employed to screen the entire open reading frame of the gene.

Various modifications and variations of the described compositions, methods, and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific

embodiments. Nor should the description of such embodiments be considered exhaustive or limit the invention to the precise forms disclosed. Furthermore, elements from one embodiment can be readily recombined with elements from one or more other embodiments. Such combinations can form a number of embodiments within the scope of the invention. It is intended that the scope of the invention be defined by the following claims and their equivalents.

TABLE 1

```
AACT
Full name : alpha-1-antichymotrypsin
Link : AACT_link_cdna
  Subsequence      GB:AACT      1      1520   #1
  CDS  GB:AACT.1    1302 bp            #2
      ORF    12     1313
  Allele        GB:AACT       1      36     36     A>G
                source        isSNP  SNP00027203
                consequence   GB:AACT.1     2      Missense      9-9    A>T
  Allele        GB:AACT       1      269    269    A>G
                source        isSNP  SNP00073834
                consequence   GB:AACT.1     2      Silent        86-86  F
  Allele        GB:AACT       1      830    830    A>G
                source        isSNP  SNP00047132
                consequence   GB:AACT.1     2      Silent        273-273      S
  Allele        GB:AACT       1      836    836    A>G
                source        isSNP  SNP00043844
                consequence   GB:AACT.1     2      Silent        275-275      L
  Allele        GB:AACT       1      837    837    A>G
                source        isSNP  SNP00101207
                consequence   GB:AACT.1     2      Missense      276-276      F>L
  Allele        GB:AACT       1      848    848    A>G
                source        isSNP  SNP00101208
                consequence   GB:AACT.1     2      Silent        279-279      P
  Allele        GB:AACT       1      854    854    A>G
                source        isSNP  SNP00052361
                consequence   GB:AACT.1     2      Silent        281-281      Q
  Allele        GB:AACT       1      947    947    G>T
                source        isSNP  SNP00059862
                consequence   GB:AACT.1     2      Stop   312-312
  Allele        GB:AACT       1      1227   1227   A>G
                source        isSNP  SNP00046872
                consequence   GB:AACT.1     2      Missense      406-406      T>A
  GIF AACT-cdna-fwd.gif
Link : FL_2114865_link_genomic
  Subsequence      GB:AL049839_2        1      214520      #3
  Subsequence      AACT_mrna_build.1  59531  69154  #4
  Subsequence      AACT_cds.2    59542  67448  #5
  CDS  AACT_cds.2    651 bp       2 exons       #5
       exon   59542  60184
       exon   67441  67448
  mRNA        AACT_mrna_build.1  1523 bp      4 exons       #4
       exon   59531  60184
       exon   64295  64568
       exon   67441  67591
       exon   68711  69154
  Allele       GB:AL049839_2       3      59566  59566  A>G
               source        isSNP  SNP00027203
               source        wetSNP        GB:AL049839_2.v59566.G>A
               consequence   AACT_cds.2    5      Missense      9-9    A>T
  Allele       GB:AL049839_2       3      59799  59799  A>G
               source        isSNP  SNP00073834
               consequence   AACT_cds.2    5      Silent        86-86  F
  Allele       GB:AL049839_2       3      59844  59844  A>G
```

TABLE 1 (Cont.)

```
            source         isSNP SNP00005018
            consequence  AACT_cds.2    5      Silent        101-101      K
Allele      GB:AL049839_2        3        60144  60144  A>G
            source         isSNP SNP00093217
            consequence  AACT_cds.2    5      Silent        201-201      S
Allele      GB:AL049839_2        3        64470  64470  A>G
            source         isSNP SNP00047132
            consequence  AACT_cds.2    5      Intron
Allele      GB:AL049839_2        3        64476  64476  A>G
            source         isSNP SNP00043844
            consequence  AACT_cds.2    5      Intron
Allele      GB:AL049839_2        3        64477  64477  A>G
            source         isSNP SNP00101207
            consequence  AACT_cds.2    5      Intron
Allele      GB:AL049839_2        3        64488  64488  A>G
            source         isSNP SNP00101208
            consequence  AACT_cds.2    5      Intron
Allele      GB:AL049839_2        3        64494  64494  A>G
            source         isSNP SNP00052361
            consequence  AACT_cds.2    5      Intron
Allele      GB:AL049839_2        3        65434  65434  A>G
            source         isSNP SNP00052361
            consequence  AACT_cds.2    5      Intron
Allele      GB:AL049839_2        3        65440  65440  A>G
            source         isSNP SNP00101208
            consequence  AACT_cds.2    5      Intron
Allele      GB:AL049839_2        3        65451  65451  A>G
            source         isSNP SNP00101207
            consequence  AACT_cds.2    5      Intron
Allele      GB:AL049839_2        3        65452  65452  A>G
            source         isSNP SNP00043844
            consequence  AACT_cds.2    5      Intron
Allele      GB:AL049839_2        3        65458  65458  A>G
            source         isSNP SNP00047132
            consequence  AACT_cds.2    5      Intron
Allele      GB:AL049839_2        3        68858  68858  A>G
            source         isSNP SNP00046872
            consequence  AACT_cds.2    5      3'
Allele      GB:AL049839_2        3        68882  68882  A>G
            source         wetSNP         GB:AL049839_2.v68882.A>G
            consequence  AACT_cds.2    5      3'
 GIF AACT-genomic-fwd.gif


ABL1
Full name : v-abl Abelson murine leukemia viral oncogene homolog 1
Link : ABL1_link_cdna
  Subsequence        GB:NM_005157        1       5744    #6
  CDS GB:NM_005157.1       3393 bp               #7
     ORF   365    3757
  Allele      GB:NM_005157         6       1916   1916   C>G
              source         isSNP SNP00046020
              consequence  GB:NM_005157.1      7       Missense      518-518      A>P
  Allele      GB:NM_005157         6       2716   2716   C>G
```

TABLE 1 (Cont.)

```
                source          isSNP  SNP00068702
                consequence  GB:NM_005157.1      7      Silent        784-784     P
   Allele       GB:NM_005157          6      3625   3625   A>G
                source          isSNP  SNP00098956
                consequence  GB:NM_005157.1      7      Silent        1087-1087   I
   Allele       GB:NM_005157          6      3688   3688   A>G
                source          isSNP  SNP00012765
                consequence  GB:NM_005157.1      7      Silent        1108-1108   P
   Allele       GB:NM_005157          6      3894   3894   C>G
                source          isSNP  SNP00046021
                consequence  GB:NM_005157.1      7      3'
   Allele       GB:NM_005157          6      4612   4612   A>G
                source          isSNP  SNP00051628
                consequence  GB:NM_005157.1      7      3'
   Allele       GB:NM_005157          6      5512   5512   A>G
                source          isSNP  SNP00012768
                consequence  GB:NM_005157.1      7      3'
   GIF  ABL1-cdna-fwd.gif
 Link : ABL1_link_genomic
   Subsequence        ABL1_cds.1     73887  116507        #8
   Subsequence        ABL1_cds.2     29132  116507        #9
   Subsequence        GB:U07561_1    1      35962  #10
   Subsequence        GB:U07563_1    36063  120601        #11
   Subsequence        ABL1_mrna_build.1   73506  118495        #12
   Subsequence        ABL1_mrna_build.2   28792  116507        #13
   Subsequence        ABL1_mrna_build.3   73724  116507        #14
   CDS  ABL1_cds.1    3393 bp      11 exons      #8
        exon   73887  73965
        exon   85951  86124
        exon   86688  86983
        exon   94650  94922
        exon   104016        104100
        exon   104747        104924
        exon   106755        106939
        exon   109237        109389
        exon   110890        110979
        exon   111322        111486
        exon   114793        116507
   CDS  ABL1_cds.2    3450 bp      11 exons      #9
        exon   29132  29267
        exon   85951  86124
        exon   86688  86983
        exon   94650  94922
        exon   104016        104100
        exon   104747        104924
        exon   106755        106939
        exon   109237        109389
        exon   110890        110979
        exon   111322        111486
        exon   114793        116507
   mRNA        ABL1_mrna_build.1  5762 bp      11 exons      #12
        exon   73506  73965
        exon   85951  86124
        exon   86688  86983
        exon   94650  94922
```

TABLE 1 (Cont.)

```
    exon   104016         104100
    exon   104747         104924
    exon   106755         106939
    exon   109237         109389
    exon   110890         110979
    exon   111322         111486
    exon   114793         118495
mRNA       ABL1_mrna_build.2  3787 bp      11 exons     #13
    exon   28792  29267
    exon   85954  86124
    exon   86688  86983
    exon   94650  94922
    exon   104016         104100
    exon   104747         104924
    exon   106755         106939
    exon   109237         109389
    exon   110890         110979
    exon   111322         111486
    exon   114793         116507
mRNA       ABL1_mrna_build.3  3556 bp      11 exons     #14
    exon   73724  73965
    exon   85951  86124
    exon   86688  86983
    exon   94650  94922
    exon   104016         104100
    exon   104747         104924
    exon   106755         106939
    exon   109237         109389
    exon   110890         110979
    exon   111322         111486
    exon   114793         116507
Allele     GB:U07561_1  10      29061  29061  A>G
           source       isSNP  SNP00120072
           consequence  ABL1_cds.1    8       5'
           consequence  ABL1_cds.2    9       5'
Allele     GB:U07561_1  10      30837  30837  A>G
           source       dbSNP  gnl|dbSNP|ss642659_allele
           source       dbSNP  gnl|dbSNP|ss1045108_allele
           source       dbSNP  gnl|dbSNP|ss1044696_allele
           consequence  ABL1_cds.1    8       5'
           consequence  ABL1_cds.2    9       Intron
Allele     GB:U07563_1  11      35864  35864  A>G
           source       isSNP  SNP00048470
           consequence  ABL1_cds.1    8       5'
           consequence  ABL1_cds.2    9       Intron
Allele     GB:U07563_1  11      58876  58876  C>G
           source       wetSNP        GB:U07563_1.v58876.C>G
           consequence  ABL1_cds.1    8       Intron
           consequence  ABL1_cds.2    9       Intron
Allele     GB:U07563_1  11      68640  68640  A>G
           source       wetSNP        GB:U07563_1.v68640.T>C
           consequence  ABL1_cds.1    8       Intron
           consequence  ABL1_cds.2    9       Intron
Allele     GB:U07563_1  11      74901  74901  A>G
           source       wetSNP        GB:U07563_1.v74901.A>G
```

TABLE 1 (Cont.)

	consequence	ABL1_cds.1	8	Silent	499-499	E	
	consequence	ABL1_cds.2	9	Silent	518-518	E	
Allele	GB:U07563_1	11	75298	75298	C>G		
	source	isSNP	SNP00046020				
	consequence	ABL1_cds.1	8	Missense	518-518	A>P	
	consequence	ABL1_cds.2	9	Missense	537-537	A>P	
Allele	GB:U07563_1	11	78921	78921	A>G		
	source	wetSNP	GB:U07563_1.v78921.G>A				
	consequence	ABL1_cds.1	8	Silent	623-623	E	
	consequence	ABL1_cds.2	9	Silent	642-642	E	
Allele	GB:U07563_1	11	79239	79239	A>G		
	source	wetSNP	GB:U07563_1.v79239.G>A				
	consequence	ABL1_cds.1	8	Silent	729-729	T	
	consequence	ABL1_cds.2	9	Silent	748-748	T	
Allele	GB:U07563_1	11	79404	79404	C>G		
	source	isSNP	SNP00068702				
	source	wetSNP	GB:U07563_1.v79404.C>G				
	consequence	ABL1_cds.1	8	Silent	784-784	P	
	consequence	ABL1_cds.2	9	Silent	803-803	P	
Allele	GB:U07563_1	11	79657	79657	A>G		
	source	wetSNP	GB:U07563_1.v79657.C>T				
	consequence	ABL1_cds.1	8	Missense	869-869	P>S	
	consequence	ABL1_cds.2	9	Missense	888-888	P>S	
Allele	GB:U07563_1	11	79750	79750	A>G		
	source	wetSNP	GB:U07563_1.v79750.C>T				
	consequence	ABL1_cds.1	8	Missense	900-900	P>S	
	consequence	ABL1_cds.2	9	Missense	919-919	P>S	
Allele	GB:U07563_1	11	80313	80313	A>G		
	source	isSNP	SNP00098956				
	consequence	ABL1_cds.1	8	Silent	1087-1087	I	
	consequence	ABL1_cds.2	9	Silent	1106-1106	I	
Allele	GB:U07563_1	11	80376	80376	A>G		
	source	isSNP	SNP00012765				
	source	wetSNP	GB:U07563_1.v80376.G>A				
	consequence	ABL1_cds.1	8	Silent	1108-1108	P	
	consequence	ABL1_cds.2	9	Silent	1127-1127	P	
Allele	GB:U07563_1	11	80582	80582	C>G		
	source	isSNP	SNP00046021				
	consequence	ABL1_cds.1	8	3'			
	consequence	ABL1_cds.2	9	3'			
Allele	GB:U07563_1	11	81298	81298	A>G		
	source	isSNP	SNP00051628				
	consequence	ABL1_cds.1	8	3'			
	consequence	ABL1_cds.2	9	3'			
Allele	GB:U07563_1	11	81806	81806	A>G		
	source	isSNP	SNP00012766				
	consequence	ABL1_cds.1	8	3'			
	consequence	ABL1_cds.2	9	3'			
Allele	GB:U07563_1	11	82199	82199	A>G		
	source	isSNP	SNP00012768				
	consequence	ABL1_cds.1	8	3'			
	consequence	ABL1_cds.2	9	3'			

GIF ABL1-genomic-fwd.gif

TABLE 1 (Cont.)

```
ADAM9
Full name : a disintegrin and metalloproteinase domain 9
Link : ADAM9_link_cdna
  Subsequence      GB:HSU41766  1      3865    #15
  CDS  GB:HSU41766.1        2460 bp            #16
       ORF    79     2538
  Allele      GB:HSU41766  15     462    462    G>T
              source         isSNP  SNP00060630
              consequence  GB:HSU41766.1        16     Missense      128-128      I>M
  Allele      GB:HSU41766  15     1486   1486   A>G
              source         isSNP  SNP00122821
              consequence  GB:HSU41766.1        16     Missense      470-470      G>S
  Allele      GB:HSU41766  15     1580   1580   G>T
              source         isSNP  SNP00060631
              consequence  GB:HSU41766.1        16     Missense      501-501      N>T
  Allele      GB:HSU41766  15     2845   2845   A>G
              source         isSNP  SNP00024957
              consequence  GB:HSU41766.1        16     3'
  Allele      GB:HSU41766  15     3112   3112   A>G
              source         isSNP  SNP00122822
              consequence  GB:HSU41766.1        16     3'
  Allele      GB:HSU41766  15     3703   3703   A>G
              source         isSNP  SNP00024958
              consequence  GB:HSU41766.1        16     3'
  GIF  ADAM9-cdna-fwd.gif


ADAMTS1
Full name : a disintegrin-like and metalloprotease (reprolysin  type) with
thrombospondin type 1 motif, 1
Link : ADAMTS1_link_cdna
  Subsequence      GB:AF060152_1        1      3430    #17
  CDS  GB:AF060152_1.1      2853 bp            #18
       ORF    238    3090
  Allele      GB:AF060152_1        17     140    140     C>G
              source         isSNP  SNP00109009
              consequence  GB:AF060152_1.1      18     5'
  Allele      GB:AF060152_1        17     282    282     G>T
              source         isSNP  SNP00071624
              consequence  GB:AF060152_1.1      18     Silent        15-15  P
  Allele      GB:AF060152_1        17     768    768     G>T
              source         isSNP  SNP00069180
              consequence  GB:AF060152_1.1      18     Silent        177-177      V
  Allele      GB:AF060152_1        17     865    865     C>G
              source         isSNP  SNP00069181
              consequence  GB:AF060152_1.1      18     Missense      210-210      P>A
  Allele      GB:AF060152_1        17     1686   1686    A>G
              source         isSNP  SNP00033973
              consequence  GB:AF060152_1.1      18     Silent        483-483      P
  Allele      GB:AF060152_1        17     2294   2294    A>G
              source         isSNP  SNP00109010
              consequence  GB:AF060152_1.1      18     Missense      686-686      R>H
  Allele      GB:AF060152_1        17     2370   2370    A>G
              source         isSNP  SNP00033974
```

TABLE 1 (Cont.)

```
            consequence GB:AF060152_1.1     18     Silent        711-711      S
  Allele    GB:AF060152_1       17     2958    2958   A>G
            source        isSNP  SNP00033975
            consequence GB:AF060152_1.1     18     Silent        907-907      C
  GIF  ADAMTS1-cdna-fwd.gif


ADAMTS4
Full name : a disintegrin-like and metalloprotease (reprolysin type) with
thrombospondin type 1 motif, 4
Link : ADAMTS4_link_cdna
  Subsequence      GB:NM_005099_1      1      4301   #19
  CDS GB:NM_005099_1.1   2514 bp              #20
      ORF    401    2914
  Allele     GB:NM_005099_1      19     2970   2970   A>G
             source        isSNP  SNP00022951
             consequence  GB:NM_005099_1.1    20     3'
  Allele     GB:NM_005099_1      19     3529   3529   A>G
             source        dbSNP  gnl|dbSNP|ss610462_allele
             consequence  GB:NM_005099_1.1    20     3'
  Allele     GB:NM_005099_1      19     3533   3533   A>G
             source        dbSNP  gnl|dbSNP|ss722414_allele
             source        dbSNP  gnl|dbSNP|ss999631_allele
             consequence  GB:NM_005099_1.1    20     3'
  Allele     GB:NM_005099_1      19     3855   3855   A>G
             source        dbSNP  gnl|dbSNP|ss1298908_allele
             consequence  GB:NM_005099_1.1    20     3'
  GIF  ADAMTS4-cdna-fwd.gif


AGC1
Full name : aggrecan 1
Link : AGC1_link_cdna
  Subsequence      GB:HUMAGPRO  1      7137   #21
  CDS GB:HUMAGPRO.1        6951 bp            #22
      ORF    61     7011
  Allele     GB:HUMAGPRO   21     6495   6495   G>T
             source        isSNP  SNP00010327
             consequence  GB:HUMAGPRO.1       22     Silent        2145-2145    A
  GIF  AGC1-cdna-fwd.gif


ANK
Full name : human homolog of mouse ank gene
Link : ANK_fl_link_cdna
  Subsequence      FN:3255641CB1       1      1481   #23
  CDS FN:3255641CB1.1     1338 bp             #24
      ORF    106    1443
  Allele     FN:3255641CB1       23     258    258    A>G
             source        isSNP  SNP00073561
             consequence  FN:3255641CB1.1     24     Silent        51-51  A
  Allele     FN:3255641CB1       23     1048   1048   C>G
```

TABLE 1 (Cont.)

```
            source        isSNP SNP00036339
            consequence FN:3255641CB1.1    24    Missense    315-315     A>P
  Allele    FN:3255641CB1       23    1106   1106   A>G
            source        isSNP SNP00075037
            consequence FN:3255641CB1.1    24    Missense    334-334     V>A
  Allele    FN:3255641CB1       23    1373   1373   A>G
            source        isSNP SNP00045819
            consequence FN:3255641CB1.1    24    Missense    423-423     S>F
  GIF ANK-cdna-fwd.gif
Link : ANK_link_cdna
  Subsequence      GB:AF274753_1       1      1568   #25
  CDS GB:AF274753_1.1     1479 bp             #26
      ORF    69     1547
  Allele    GB:AF274753_1       25    362    362    A>G
            source        isSNP SNP00073561
            consequence GB:AF274753_1.1    26    Silent      98-98  A
  Allele    GB:AF274753_1       25    1152   1152   C>G
            source        isSNP SNP00036339
            consequence GB:AF274753_1.1    26    Missense    362-362     A>P
  Allele    GB:AF274753_1       25    1210   1210   A>G
            source        isSNP SNP00075037
            consequence GB:AF274753_1.1    26    Missense    381-381     V>A
  Allele    GB:AF274753_1       25    1477   1477   A>G
            source        isSNP SNP00045819
            consequence GB:AF274753_1.1    26    Missense    470-470     S>F
  GIF ANK-cdna-fwd.gif
Link : ANK_link_genomic
  Subsequence      ANK_cds.1     26332  84281  #27
  Subsequence      GBI:AC016575_6_000010     1      605    #28
  Subsequence      GB:AC026437_2       706    92528  #29
  Subsequence      ANK_mrna_build.1    308    85658  #30
  Subsequence      ANK_cds.2     272    84281  #31
  CDS  ANK_cds.1     1338 bp      11 exons     #27
       exon   26332  26503
       exon   36882  37000
       exon   39535  39618
       exon   44240  44410
       exon   46173  46307
       exon   49517  49609
       exon   53557  53652
       exon   78643  78772
       exon   81811  81934
       exon   82505  82604
       exon   84168  84281
  CDS  ANK_cds.2     1479 bp      12 exons     #31
       exon   272    367
       exon   26287  26503
       exon   36882  37000
       exon   39535  39618
       exon   44240  44410
       exon   46173  46307
       exon   49517  49609
       exon   53557  53652
       exon   78643  78772
       exon   81811  81934
```

TABLE 1 (Cont.)

	exon	82505 82604					
	exon	84168 84281					
mRNA		ANK_mrna_build.1	2820 bp	12 exons	#30		
	exon	308 367					
	exon	26287 26503					
	exon	36882 37000					
	exon	39535 39618					
	exon	44240 44410					
	exon	46173 46307					
	exon	49517 49609					
	exon	53557 53652					
	exon	78643 78772					
	exon	81811 81934					
	exon	82505 82604					
	exon	84168 85658					
Allele	GB:AC026437_2		29	8413 8413	C>G		
	source	dbSNP	gnl|dbSNP|ss95678_allele				
	consequence	ANK_cds.1	27	5'			
	consequence	ANK_cds.2	31	Intron			
Allele	GB:AC026437_2		29	14825 14825	A>G		
	source	dbSNP	gnl|dbSNP|ss619053_allele				
	source	dbSNP	gnl|dbSNP|ss1002004_allele				
	source	dbSNP	gnl|dbSNP|ss227983_allele				
	source	dbSNP	gnl|dbSNP|ss324626_allele				
	consequence	ANK_cds.1	27	5'			
	consequence	ANK_cds.2	31	Intron			
Allele	GB:AC026437_2		29	25779 25779	A>G		
	source	wetSNP	GB:AC026437_2.v25779.C>T				
	consequence	ANK_cds.1	27	Silent		51-51	A
	consequence	ANK_cds.2	31	Silent		98-98	A
Allele	GB:AC026437_2		29	25807 25807	A>G		
	source	isSNP	SNP00104502				
	source	wetSNP	GB:AC026437_2.v25807.G>A				
	consequence	ANK_cds.1	27	Intron			
	consequence	ANK_cds.2	31	Intron			
Allele	GB:AC026437_2		29	26433 26433	A>G		
	source	isSNP	SNP00018441				
	consequence	ANK_cds.1	27	Intron			
	consequence	ANK_cds.2	31	Intron			
Allele	GB:AC026437_2		29	30696 30696	A>T		
	source	dbSNP	gnl|dbSNP|ss1016631_allele				
	source	dbSNP	gnl|dbSNP|ss389763_allele				
	consequence	ANK_cds.1	27	Intron			
	consequence	ANK_cds.2	31	Intron			
Allele	GB:AC026437_2		29	34277 34277	A>G		
	source	isSNP	SNP00101566				
	consequence	ANK_cds.1	27	Intron			
	consequence	ANK_cds.2	31	Intron			
Allele	GB:AC026437_2		29	36172 36172	A>G		
	source	wetSNP	GB:AC026437_2.v36172.T>C				
	consequence	ANK_cds.1	27	Intron			
	consequence	ANK_cds.2	31	Intron			
Allele	GB:AC026437_2		29	37028 37028	G>T		
	source	isSNP	SNP00056800				
	consequence	ANK_cds.1	27	Intron			

TABLE 1 (Cont.)

	consequence	ANK_cds.2		31	Intron			
Allele	GB:AC026437_2		29	37186	37186	G>T		
	source	isSNP	SNP00022144					
	consequence	ANK_cds.1		27	Intron			
	consequence	ANK_cds.2		31	Intron			
Allele	GB:AC026437_2		29	37205	37205	A>G		
	source	isSNP	SNP00022143					
	consequence	ANK_cds.1		27	Intron			
	consequence	ANK_cds.2		31	Intron			
Allele	GB:AC026437_2		29	37340	37340	A>T		
	source	dbSNP	gnl|dbSNP|ss469809_allele					
	consequence	ANK_cds.1		27	Intron			
	consequence	ANK_cds.2		31	Intron			
Allele	GB:AC026437_2		29	52817	52817	G>T		
	source	wetSNP		GB:AC026437_2.v52817.C>A				
	consequence	ANK_cds.1		27	Intron			
	consequence	ANK_cds.2		31	Intron			
Allele	GB:AC026437_2		29	52899	52899	A>G		
	source	wetSNP		GB:AC026437_2.v52899.A>G				
	consequence	ANK_cds.1		27	Silent		274-274	A
	consequence	ANK_cds.2		31	Silent		321-321	A
Allele	GB:AC026437_2		29	52962	52962	G>T		
	source	wetSNP		GB:AC026437_2.v52962.T>G				
	consequence	ANK_cds.1		27	Intron			
	consequence	ANK_cds.2		31	Intron			
Allele	GB:AC026437_2		29	63950	63950	A>G		
	source	isSNP	SNP00093702					
	consequence	ANK_cds.1		27	Intron			
	consequence	ANK_cds.2		31	Intron			
Allele	GB:AC026437_2		29	78010	78010	C>G		
	source	isSNP	SNP00036339					
	consequence	ANK_cds.1		27	Missense		315-315	A>P
	consequence	ANK_cds.2		31	Missense		362-362	A>P
Allele	GB:AC026437_2		29	78875	78875	A>G		
	source	isSNP	SNP00095793					
	consequence	ANK_cds.1		27	Intron			
	consequence	ANK_cds.2		31	Intron			
Allele	GB:AC026437_2		29	81235	81235	A>G		
	source	wetSNP		GB:AC026437_2.v81235.T>C				
	consequence	ANK_cds.1		27	Intron			
	consequence	ANK_cds.2		31	Intron			
Allele	GB:AC026437_2		29	82852	82852	A>G		
	source	isSNP	SNP00120424					
	consequence	ANK_cds.1		27	Intron			
	consequence	ANK_cds.2		31	Intron			
Allele	GB:AC026437_2		29	83057	83057	A>G		
	source	isSNP	SNP00120425					
	consequence	ANK_cds.1		27	Intron			
	consequence	ANK_cds.2		31	Intron			
Allele	GB:AC026437_2		29	83506	83506	A>G		
	source	isSNP	SNP00045819					
	consequence	ANK_cds.1		27	Missense		423-423	S>F
	consequence	ANK_cds.2		31	Missense		470-470	S>F
Allele	GB:AC026437_2		29	83587	83587	A>G		
	source	wetSNP		GB:AC026437_2.v83587.G>A				

TABLE 1 (Cont.)

```
          consequence  ANK_cds.1    27     3'
          consequence  ANK_cds.2    31     3'
  Allele    GB:AC026437_2      29     83607  83607  A>G
          source       isSNP  SNP00008779
          source       wetSNP        GB:AC026437_2.v83607.A>G
          consequence  ANK_cds.1    27     3'
          consequence  ANK_cds.2    31     3'
  Allele    GB:AC026437_2      29     84086  84086  A>G
          source       isSNP  SNP00012596
          consequence  ANK_cds.1    27     3'
          consequence  ANK_cds.2    31     3'
  Allele    GB:AC026437_2      29     84156  84156  A>G
          source       isSNP  SNP00045820
          consequence  ANK_cds.1    27     3'
          consequence  ANK_cds.2    31     3'
  Allele    GB:AC026437_2      29     84896  84896  G>T
          source       isSNP  SNP00045822
          consequence  ANK_cds.1    27     3'
          consequence  ANK_cds.2    31     3'
  GIF ANK-genomic-fwd.gif


BGLAP
Full name : Bone Gla Protein
Link : FL_104137_link_genomic
  Subsequence      GB:AC007227_2_104137CD1   35521  34594  #32
  Subsequence      GB:AC007227_2       1        167932        #33
  Subsequence      BGLAP_mrna_build.1         35539  34461  #34
  mRNA       BGLAP_mrna_build.1        451 bp       4 exons      #34
      exon  35539  35458
      exon  35200  35162
      exon  34991  34922
      exon  34720  34461
  CDS  GB:AC007227_2_104137CD1   300 bp       4 exons      #32
      exon  35521  35458
      exon  35200  35162
      exon  34991  34922
      exon  34720  34594
  Allele    GB:AC007227_2       33     34618  34618  C>G
          source       wetSNP       GB:AC007227_2.v34618.G>C
          consequence  GB:AC007227_2_104137CD1   32     Silent       92-92  A
  Allele    GB:AC007227_2       33     34977  34977  G>T
          source       wetSNP       GB:AC007227_2.v34977.G>T
          consequence  GB:AC007227_2_104137CD1   32     Missense     40-40  Q>K
  Allele    GB:AC007227_2       33     35228  35228  C>G
          source       isSNP  SNP00038471
          consequence  GB:AC007227_2_104137CD1   32     Intron
  GIF BGLAP-genomic-rev.gif


BGN
Full name : BGN
Link : BGN_link_cdna
```

TABLE 1 (Cont.)

```
  Subsequence      GB:HUMHPGI   1      1685   #35
  CDS GB:HUMHPGI.1        1107 bp             #36
      ORF    121    1227
  Allele      GB:HUMHPGI   35     70     70     G>T
              source       isSNP  SNP00011488
              consequence  GB:HUMHPGI.1         36     5'
  Allele      GB:HUMHPGI   35     261    261    A>G
              source       isSNP  SNP00011489
              consequence  GB:HUMHPGI.1         36     Silent        47-47  S
  Allele      GB:HUMHPGI   35     660    660    A>G
              source       isSNP  SNP00011490
              consequence  GB:HUMHPGI.1         36     Silent        180-180      S
  Allele      GB:HUMHPGI   35     1355   1355   A>G
              source       isSNP  SNP00092805
              consequence  GB:HUMHPGI.1         36     3'
  GIF BGN-cdna-fwd.gif
Link : BGN_link_genomic
  Subsequence      GB:U82695    1      76146  #37
  Subsequence      GB:U82695_2540367CD1       18042  21854  #38
  Subsequence      BGN_mrna_build.1    8415   22311  #39
  CDS GB:U82695_2540367CD1      1107 bp       7 exons       #38
      exon   18042  18279
      exon   18648  18760
      exon   19272  19485
      exon   19938  20048
      exon   20239  20332
      exon   20456  20594
      exon   21657  21854
  mRNA        BGN_mrna_build.1   1684 bp      8 exons       #39
      exon   8415   8523
      exon   18031  18279
      exon   18648  18760
      exon   19272  19485
      exon   19938  20048
      exon   20239  20332
      exon   20456  20594
      exon   21657  22311
  Allele      GB:U82695    37     8484   8484   G>T
              source       isSNP  SNP00011488
              consequence  GB:U82695_2540367CD1     38     5'
  Allele      GB:U82695    37     18161  18161  A>G
              source       wetSNP        GB:U82695.v18161.A>G
              consequence  GB:U82695_2540367CD1     38     Silent        40-40  E
  Allele      GB:U82695    37     18182  18182  A>G
              source       isSNP  SNP00011489
              source       wetSNP        GB:U82695.v18182.G>A
              consequence  GB:U82695_2540367CD1     38     Silent        47-47  S
  Allele      GB:U82695    37     18330  18330  A>G
              source       wetSNP        GB:U82695.v18330.G>A
              consequence  GB:U82695_2540367CD1     38     Intron
  Allele      GB:U82695    37     18354  18354  A>G
              source       wetSNP        GB:U82695.v18354.G>A
              consequence  GB:U82695_2540367CD1     38     Intron
  Allele      GB:U82695    37     19460  19460  A>G
              source       isSNP  SNP00011490
```

TABLE 1 (Cont.)

```
            source       wetSNP      GB:U82695.v19460.T>C
            consequence  GB:U82695_2540367CD1     38     Silent       180-180
    S
  Allele    GB:U82695    37     21566  21566  G>T
            source       wetSNP      GB:U82695.v21566.G>T
            consequence  GB:U82695_2540367CD1     38     Intron
  Allele    GB:U82695    37     21639  21639  A>G
            source       wetSNP      GB:U82695.v21639.C>T
            consequence  GB:U82695_2540367CD1     38     Intron
  Allele    GB:U82695    37     21982  21982  A>G
            source       isSNP  SNP00092805
            consequence  GB:U82695_2540367CD1     38     3'
  Allele    GB:U82695    37     22172  22172  G>T
            source       isSNP  SNP00011491
            consequence  GB:U82695_2540367CD1     38     3'
  GIF BGN-genomic-fwd.gif


BHLHB2
Full name : basic helix-loop-helix domain containing, class B, 2
Link : BHLHB2_link_cdna
  Subsequence       GB:AB004066_1      1      2922   #40
  CDS  GB:AB004066_1.1    1239 bp             #41
      ORF    197    1435
  Allele      GB:AB004066_1      40     196     196    A>G
              source       isSNP  SNP00062724
              consequence  GB:AB004066_1.1    41     5'
  Allele      GB:AB004066_1      40     829     829    A>G
              source       isSNP  SNP00046376
              consequence  GB:AB004066_1.1    41     Silent       211-211      G
  Allele      GB:AB004066_1      40     2070    2070   A>G
              source       isSNP  SNP00013041
              consequence  GB:AB004066_1.1    41     3'
  Allele      GB:AB004066_1      40     2323    2323   A>G
              source       isSNP  SNP00013042
              consequence  GB:AB004066_1.1    41     3'
  GIF  BHLHB2-cdna-fwd.gif


BMP2
Full name : BMP2
Link : BMP2_link_cdna
  Subsequence       GB:HUMBMP2A  1      1547   #42
  CDS  GB:HUMBMP2A.1      1191 bp             #43
      ORF    324    1514
  Allele      GB:HUMBMP2A  42     584    584    A>G
              source       isSNP  SNP00015730
              consequence  GB:HUMBMP2A.1      43     Silent       87-87  S
  Allele      GB:HUMBMP2A  42     760    760    A>G
              source       isSNP  SNP00015731
              consequence  GB:HUMBMP2A.1      43     Missense     146-146      T>I
  Allele      GB:HUMBMP2A  42     984    984    G>T
              source       isSNP  SNP00015732
```

TABLE 1 (Cont.)

```
            consequence  GB:HUMBMP2A.1      43    Missense    221-221     H>N
  Allele    GB:HUMBMP2A  42     1484   1484   A>G
            source       isSNP  SNP00015733
            consequence  GB:HUMBMP2A.1      43    Silent      387-387     D
  GIF  BMP2-cdna-fwd.gif
Link : FL_3220019_link_genomic
  Subsequence       GB:HS859D4  1       178870       #44
  Subsequence       GB:HS859D4_3220019CD1   176685       167723       #45
  Subsequence       BMP2_mrna_build.1  178252       167687       #46
  mRNA      BMP2_mrna_build.1  1547 bp      3 exons      #46
      exon  178252        177937
      exon  176692        176340
      exon  168564        167687
  CDS  GB:HS859D4_3220019CD1     1188 bp      2 exons      #45
      exon  176685        176340
      exon  168564        167723
  Allele    GB:HS859D4   44    167750       167750       A>G
            source       isSNP  SNP00015733
            consequence  GB:HS859D4_3220019CD1     45    Silent      387-387
      D
  Allele    GB:HS859D4   44    168250       168250       G>T
            source       isSNP  SNP00015732
            consequence  GB:HS859D4_3220019CD1     45    Missense    221-221
      H>N
  Allele    GB:HS859D4   44    168341       168341       A>T
            source       wetSNP      GB:HS859D4.v168341.T>A
            consequence  GB:HS859D4_3220019CD1     45    Missense    190-190
      R>S
  Allele    GB:HS859D4   44    168474       168474       A>G
            source       isSNP  SNP00015731
            consequence  GB:HS859D4_3220019CD1     45    Missense    146-146
      T>I
  Allele    GB:HS859D4   44    176425       176425       A>G
            source       isSNP  SNP00015730
            source       wetSNP      GB:HS859D4.v176425.T>C
            consequence  GB:HS859D4_3220019CD1     45    Silent      87-87  S
  GIF  BMP2-genomic-rev.gif


BMP4
Full name : BMP4
Link : BMP4_link_cdna
  Subsequence       GB:HUMBMP2B  1      1751   #47
  CDS  GB:HUMBMP2B.1       1227 bp            #48
      ORF   395    1621
  Allele    GB:HUMBMP2B  47     308    308    A>G
            source       isSNP  SNP00074676
            consequence  GB:HUMBMP2B.1      48     5'
  Allele    GB:HUMBMP2B  47     849    849    A>G
            source       isSNP  SNP00000573
            consequence  GB:HUMBMP2B.1      48    Missense     152-152     V>A
  GIF  BMP4-cdna-fwd.gif
 Link : BMP4_link_genomic
  Subsequence       GB:HSU43842  1      11233  #49
```

TABLE 1 (Cont.)

```
  Subsequence      GB:HSU43842_1613615CD1    7798   9984   #50
  Subsequence      BMP4_mrna_build.1  3207   10117  #51
  mRNA        BMP4_mrna_build.1  1751 bp      4 exons      #51
      exon    3207   3468
      exon    6620   6744
      exon    7791   8167
      exon    9131   10117
  CDS  GB:HSU43842_1613615CD1    1224 bp      2 exons      #50
      exon    7798   8167
      exon    9131   9984
  Allele      GB:HSU43842  49     6665   6665   A>G
              source       isSNP  SNP00074676
              consequence  GB:HSU43842_1613615CD1    50     5'
  Allele      GB:HSU43842  49     7752   7752   A>G
              source       isSNP  SNP00117542
              consequence  GB:HSU43842_1613615CD1    50     5'
  Allele      GB:HSU43842  49     9215   9215   A>G
              source       isSNP  SNP00000573
              source       wetSNP       GB:HSU43842.v9215.C>T
              consequence  GB:HSU43842_1613615CD1    50     Missense     152-152
      A>V
  GIF  BMP4-genomic-fwd.gif


BMP6
Full name : BMP6
Link : BMP6_link_cdna
  Subsequence      GB:HUMTGFBC  1      2923   #52
  CDS  GB:HUMTGFBC.1       1542 bp            #53
      ORF     160    1701
  Allele      GB:HUMTGFBC  52     1263   1263   C>G
              source       isSNP  SNP00069306
              consequence  GB:HUMTGFBC.1       53     Silent       368-368      V
  Allele      GB:HUMTGFBC  52     2280   2280   G>T
              source       isSNP  SNP00021640
              consequence  GB:HUMTGFBC.1       53     3'
  Allele      GB:HUMTGFBC  52     2436   2436   A>G
              source       isSNP  SNP00003240
              consequence  GB:HUMTGFBC.1       53     3'
  Allele      GB:HUMTGFBC  52     2574   2574   A>G
              source       isSNP  SNP00021639
              consequence  GB:HUMTGFBC.1       53     3'
  GIF  BMP6-cdna-fwd.gif


CAPN4
Full name : calpain, small polypeptide
Link : FL_508926_link_genomic
  Subsequence      GB:CH19F24590       1      41369  #54
  Subsequence      GB:CH19F24590_3639962CD1       31006  39830  #55
  Subsequence      FL_3639962_mrna_build.1  30073  40241  #56
  Subsequence      CAPN4_cds.1  31006  39833  #57
  mRNA        FL_3639962_mrna_build.1  1309 bp      11 exons      #56
```

TABLE 1 (Cont.)

```
      exon   30073  30151
      exon   30991  31214
      exon   32294  32327
      exon   32646  32735
      exon   32903  32960
      exon   33058  33122
      exon   35800  35868
      exon   35970  36048
      exon   36190  36306
      exon   39572  39630
      exon   39807  40241
  CDS  CAPN4_cds.1  717 bp       9 exons      #57
      exon   31006  31214
      exon   32294  32327
      exon   32903  32960
      exon   33058  33122
      exon   35800  35868
      exon   35970  36048
      exon   36190  36306
      exon   39572  39630
      exon   39807  39833
  CDS  GB:CH19F24590_3639962CD1        804 bp        10 exons      #55
      exon   31006  31214
      exon   32294  32327
      exon   32646  32735
      exon   32903  32960
      exon   33058  33122
      exon   35800  35868
      exon   35970  36048
      exon   36190  36306
      exon   39572  39630
      exon   39807  39830
  GIF  CAPN4-genomic-fwd.gif


CBFA1
Full name : CBFA1
Link : CBFA1_link_cdna
  Subsequence       GB:HUMCBFA   1       1411    #58
  CDS GB:HUMCBFA.2          1323 bp              #59
      ORF    1      1323
  Allele       GB:HUMCBFA    58     260    260    A>G
               source        isSNP  SNP00063798
               consequence   GB:HUMCBFA.2         59     Missense     87-87  G>E
  GIF CBFA1-cdna-fwd.gif
Link : CBFA1_link_genomic
  Subsequence       GB:HSCBFA1S1       1      93     #60
  Subsequence       GB:HSCBFA1S2       194    669    #61
  Subsequence       GB:HSCBFA1S3       770    1034   #62
  Subsequence       GB:HSCBFA1S4       1135   1381   #63
  Subsequence       GB:HSCBFA1S5       1482   1759   #64
  Subsequence       GB:HSCBFA1S6       1860   2081   #65
  Subsequence       GB:HSCBFA1S7       2182   2301   #66
  Subsequence       GB:HSCBFA1S8       2402   3033   #67
```

TABLE 1 (Cont.)

```
  Subsequence      CBFA1_cds.1  28     2948    #68
  CDS  CBFA1_cds.1  1566 bp     8 exons        #68
       exon  28     85
       exon  261    625
       exon  821    977
       exon  1198   1302
       exon  1533   1706
       exon  1881   2042
       exon  2201   2266
       exon  2470   2948
  Allele    GB:HSCBFA1S3       62     177     177     A>G
            source       wetSNP       GB:HSCBFA1S3.v177.C>T
            consequence  CBFA1_cds.1  68      Silent        183-183      N
  Allele    GB:HSCBFA1S8       67     490     490     A>G
            source       wetSNP       GB:HSCBFA1S8.v490.C>T
            consequence  CBFA1_cds.1  68      Silent        503-503      S
  GIF  CBFA1-genomic-fwd.gif


CD36
Full name : CD36 Glycoprotein
Link : CD36_link_cdna
  Subsequence      EM:HSCD3621  1      2216    #69
  Allele      EM:HSCD3621  69     123     123     G>T
              source       isSNP  SNP00011023
  Allele      EM:HSCD3621  69     196     196     A>G
              source       isSNP  SNP00096573
  Allele      EM:HSCD3621  69     230     230     C>G
              source       isSNP  SNP00110263
  Allele      EM:HSCD3621  69     827     827     A>G
              source       isSNP  SNP00115780
  Allele      EM:HSCD3621  69     1332    1332    A>G
              source       isSNP  SNP00096574
Link : CD36_link_genomic
  Subsequence      CD36_link_cds.1     2094    6548    #70
  Subsequence      EM:HSCD36G1  101    236     #71
  Subsequence      EM:HSCD36A   338    2898    #72
  Subsequence      EM:HSCD36G4  3000   3220    #73
  Subsequence      EM:HSCD36G5  3322   3529    #74
  Subsequence      EM:HSCD36AA  3631   3999    #75
  Subsequence      EM:HSCD36G7  4101   4252    #76
  Subsequence      EM:HSCD36G8  4354   4460    #77
  Subsequence      EM:HSCD36G9  4562   4691    #78
  Subsequence      EM:HSCD36G10        4793    5042    #79
  Subsequence      EM:B74110    5144   5803    #80
  Subsequence      EM:HSCD36G12        5905    6038    #81
  Subsequence      EM:HSCD36G13        6140    6252    #82
  Subsequence      EM:HSCD36G14        6354    6847    #83
  Subsequence      EM:HSCD36G15        6949    7632    #84
  Subsequence      CD36_mrna_build.1   136     7602    #85
  mRNA        CD36_mrna_build.1  2217 bp     16 exons      #85
       exon  136    206
       exon  1446   1539
       exon  2005   2213
```

TABLE 1 (Cont.)

	exon	3030	3190			
	exon	3352	3499			
	exon	3719	3898			
	exon	4131	4222			
	exon	4384	4430			
	exon	4592	4661			
	exon	4824	5011			
	exon	5265	5383			
	exon	5935	6008			
	exon	6168	6222			
	exon	6384	6548			
	exon	6979	7071			
	exon	7152	7602			
CDS	CD36_link_cds.1	1419 bp	12 exons	#70		
	exon	2094	2213			
	exon	3030	3190			
	exon	3352	3499			
	exon	3719	3898			
	exon	4131	4222			
	exon	4384	4430			
	exon	4592	4661			
	exon	4824	5011			
	exon	5265	5383			
	exon	5935	6008			
	exon	6168	6222			
	exon	6384	6548			
Allele	EM:HSCD36A	72	1160	1160	G>T	
	source	isSNP	SNP00011023			
	consequence	CD36_link_cds.1	70	5′		
Allele	EM:HSCD36A	72	1698	1698	A>G	
	source	isSNP	SNP00096573			
	consequence	CD36_link_cds.1	70	5′		
Allele	EM:HSCD36A	72	1732	1732	C>G	
	source	isSNP	SNP00110263			
	consequence	CD36_link_cds.1	70	5′		
Allele	EM:HSCD36G4	73	102	102	C>G	
	source	wetSNP	EM:HSCD36G4.v102.G>C			
	consequence	CD36_link_cds.1	70	Missense	64-64	Q>H
Allele	EM:HSCD36AA	75	232	232	A>G	
	source	isSNP	SNP00115780			
	consequence	CD36_link_cds.1	70	Silent	191-191	P
Allele	EM:HSCD36G10	79	92	92	A>G	
	source	wetSNP	EM:HSCD36G10.v92.T>C			
	consequence	CD36_link_cds.1	70	Silent	293-293	F
Allele	EM:B74110	80	193	193	A>G	
	source	isSNP	SNP00096574			
	consequence	CD36_link_cds.1	70	Silent	360-360	L
Allele	EM:HSCD36G14	83	198	203	AAGTAT>AT	
	source	wetSNP	EM:HSCD36G14.v198.AAGTAT>AT			
	consequence	CD36_link_cds.1	70	3′		
Allele	EM:HSCD36G14	83	421	421	A>G	
	source	isSNP	SNP00041723			
	consequence	CD36_link_cds.1	70	3′		
GIF	CD36-genomic-fwd.gif					

TABLE 1 (Cont.)

```
CD68
Full name : CD68 antigen
Link : FL_3777141_link_cdna
  Subsequence      FN:3777141CB1      1      1558    #86
  CDS  FN:3777141CB1.1    1065 bp          #87
       ORF    75     1139
  Allele        FN:3777141CB1       86      834     834     G>T
                source         isSNP  SNP00006442
                consequence  FN:3777141CB1.1     87     Missense      254-254       Q>K
  Allele        FN:3777141CB1       86      1394    1394    G>T
                source         dbSNP  gnl|dbSNP|ss450666_allele
                consequence  FN:3777141CB1.1     87     3'
  Allele        FN:3777141CB1       86      1475    1475    G>T
                source         isSNP  SNP00108664
                consequence  FN:3777141CB1.1     87     3'
  GIF  CD68-cdna-fwd.gif
Link : FL_1803929_link_genomic
  Subsequence      GB:AC007421_12        1       95240   #88
  Subsequence      GB:AC007421_12_3777141CD1        92493  90660  #89
  Subsequence      FL_3777141_mrna_build.1   92567  90242  #90
  mRNA       FL_3777141_mrna_build.1   1557 bp       6  exons       #90
       exon   92567  92445
       exon   92361  91844
       exon   91705  91586
       exon   91460  91388
       exon   91275  91105
       exon   90793  90242
  CDS  GB:AC007421_12_3777141CD1        1065 bp       6  exons       #89
       exon   92493  92445
       exon   92361  91844
       exon   91705  91586
       exon   91460  91388
       exon   91275  91105
       exon   90793  90660
  Allele        GB:AC007421_12       88       90404  90404  G>T
                source         dbSNP  gnl|dbSNP|ss450666_allele
                consequence  GB:AC007421_12_3777141CD1        89      3'
  Allele        GB:AC007421_12       88       90707  90707  A>G
                source         wetSNP         GB:AC007421_12.v90707.C>T
                consequence  GB:AC007421_12_3777141CD1        89      Missense
 340-340        A>T
  Allele        GB:AC007421_12       88       91388  91388  G>T
                source         wetSNP         GB:AC007421_12.v91388.G>T
                consequence  GB:AC007421_12_3777141CD1        89      Missense
 254-254        Q>K
  Allele        GB:AC007421_12       88       92357  92357  A>G
                source         wetSNP         GB:AC007421_12.v92357.C>T
                consequence  GB:AC007421_12_3777141CD1        89      Silent
 18-18  Q
  GIF  CD68-genomic-rev.gif
```

CDO1

TABLE 1 (Cont.)

```
Full name : cysteine dioxygenase type I
Link : CDO1_link_cdna
  Subsequence      GB:HHSCYSDIO       1      1556   #91
  CDS  GB:HHSCYSDIO.1     603 bp             #92
      ORF    255    857
  Allele      GB:HHSCYSDIO        91     100    100    A>G
              source       isSNP SNP00009024
              consequence  GB:HHSCYSDIO.1     92     5'
  Allele      GB:HHSCYSDIO        91     737    737    A>G
              source       isSNP SNP00048574
              consequence  GB:HHSCYSDIO.1     92     Silent      161-161      F
  Allele      GB:HHSCYSDIO        91     784    784    A>G
              source       isSNP SNP00036859
              consequence  GB:HHSCYSDIO.1     92     Missense    177-177      V>A
  Allele      GB:HHSCYSDIO        91     1082   1082   A>G
              source       isSNP SNP00107326
              consequence  GB:HHSCYSDIO.1     92     3'
  Allele      GB:HHSCYSDIO        91     1525   1525   A>G
              source       isSNP SNP00036860
              consequence  GB:HHSCYSDIO.1     92     3'
  GIF CDO1-cdna-fwd.gif
Link : CDO1_link_genomic
  Subsequence      CDO1_cds.1   1653   4275   #93
  Subsequence      GB:D85778_1  1      2601   #94
  Subsequence      GB:D85779_1  2702   2938   #95
  Subsequence      GB:D85780_1  3039   3525   #96
  Subsequence      GB:D85781_1  3626   4090   #97
  Subsequence      GB:D85782_1  4191   4921   #98
  Subsequence      CDO1_mrna_build.1  1402   4921   #99
  mRNA        CDO1_mrna_build.1  1500 bp      5 exons      #99
       exon   1402   1822
       exon   2789   2866
       exon   3178   3332
       exon   3777   3946
       exon   4246   4921
  CDS  CDO1_cds.1    603 bp       5 exons       #93
       exon   1653   1822
       exon   2789   2866
       exon   3178   3332
       exon   3777   3946
       exon   4246   4275
  Allele      GB:D85778_1  94     1498   1498   A>G
              source       isSNP SNP00009024
              consequence  CDO1_cds.1    93     5'
  Allele      GB:D85781_1  97     278    278    A>G
              source       isSNP SNP00036859
              consequence  CDO1_cds.1    93     Missense     177-177      V>A
  Allele      GB:D85782_1  98     310    310    A>G
              source       isSNP SNP00107326
              consequence  CDO1_cds.1    93     3'
  GIF CDO1-genomic-fwd.gif
```

TABLE 1 (Cont.)

```
Link : CGI-52_link_cdna
  Subsequence      GB:AF151810  1      1414   #100
  CDS  GB:AF151810.1      1080 bp            #101
      ORF    277    1356
  Allele     GB:AF151810  100    1335   1335   A>G
             source       isSNP  SNP00054191
             consequence  GB:AF151810.1      101    Silent       353-353      D
  GIF  CGI-52-cdna-fwd.gif
Link : CGI-52_link_genomic
  Subsequence      GB:AC023176_7       1      193672      #102
  Subsequence      CGI-52_mrna_build.1        131456      93050  #103
  mRNA       CGI-52_mrna_build.1       1420 bp       7 exons      #103
      exon   131456       131084
      exon   119505       119186
      exon   97592  97445
      exon   96844  96741
      exon   96095  95978
      exon   93964  93912
      exon   93353  93050
  Allele     GB:AC023176_7       102    93129  93129  A>G
             source       isSNP  SNP00054191
  Allele     GB:AC023176_7       102    93416  93416  A>G
             source       isSNP  SNP00057212
  Allele     GB:AC023176_7       102    131305       131305       C>G
             source       isSNP  SNP00069496
  GIF  CGI-52-genomic-rev.gif


CHI3L1
Full name : chitinase 3-like 1
Link : CHI3L1_link_cdna
  Subsequence      GB:NM_001276_1      1      1925   #104
  CDS  GB:NM_001276_1.1   1152 bp            #105
      ORF    127    1278
  Allele     GB:NM_001276_1      104    559    559    A>G
             source       isSNP  SNP00008252
             consequence  GB:NM_001276_1.1   105    Missense     145-145      R>G
  Allele     GB:NM_001276_1      104    590    590    A>G
             source       isSNP  SNP00071935
             consequence  GB:NM_001276_1.1   105    Missense     155-155      K>R
  Allele     GB:NM_001276_1      104    646    646    G>T
             source       isSNP  SNP00022932
             consequence  GB:NM_001276_1.1   105    Missense     174-174      L>I
  Allele     GB:NM_001276_1      104    1300   1300   A>G
             source       isSNP  SNP00052666
             consequence  GB:NM_001276_1.1   105    3'
  Allele     GB:NM_001276_1      104    1342   1342   A>G
             source       isSNP  SNP00072805
             consequence  GB:NM_001276_1.1   105    3'
  Allele     GB:NM_001276_1      104    1739   1739   A>G
             source       isSNP  SNP00076686
             consequence  GB:NM_001276_1.1   105    3'
  GIF  CHI3L1-cdna-fwd.gif
Link : CHI3L1_link_genomic
```

TABLE 1 (Cont.)

```
Subsequence       CHI3L1_cds.1        1295   7276   #106
Subsequence       CHI3L1_cds.2        1295   7433   #107
Subsequence       CHI3L1_cds.3        1295   7276   #108
Subsequence       CHI3L1_cds.4        1295   2802   #109
Subsequence       GB:Y08374_1  1      1635   #110
Subsequence       GB:Y08375_1  1736   3186   #111
Subsequence       GB:Y08376_1  3287   4116   #112
Subsequence       GB:Y08377_1  4217   5035   #113
Subsequence       GB:Y08378_1  5136   7923   #114
Subsequence       CHI3L1_mrna_build.1        1169   7923   #115
Subsequence       CHI3L1_mrna_build.2        1169   7604   #116
mRNA       CHI3L1_mrna_build.2       1355 bp      11 exons      #116
     exon   1169   1319
     exon   1572   1601
     exon   2036   2237
     exon   2789   2845
     exon   3606   3756
     exon   4517   4638
     exon   5436   5559
     exon   6069   6251
     exon   6844   6960
     exon   7296   7456
     exon   7548   7604
CDS  CHI3L1_cds.1        1152 bp      10 exons      #106
     exon   1295   1319
     exon   1572   1601
     exon   2036   2237
     exon   2789   2845
     exon   3606   3756
     exon   4517   4638
     exon   5436   5559
     exon   6069   6251
     exon   6844   6960
     exon   7136   7276
CDS  CHI3L1_cds.2        1149 bp      10 exons      #107
     exon   1295   1319
     exon   1572   1601
     exon   2036   2237
     exon   2789   2845
     exon   3606   3756
     exon   4517   4638
     exon   5436   5559
     exon   6069   6251
     exon   6844   6960
     exon   7296   7433
CDS  CHI3L1_cds.3        969 bp       9 exons       #108
     exon   1295   1319
     exon   1572   1601
     exon   2036   2237
     exon   2789   2845
     exon   3606   3756
     exon   4517   4638
     exon   5436   5559
     exon   6844   6960
     exon   7136   7276
```

TABLE 1 (Cont.)

mRNA	CHI3L1_mrna_build.1		1925 bp		10 exons	#115	
exon	1169	1319					
exon	1572	1601					
exon	2036	2237					
exon	2789	2845					
exon	3606	3756					
exon	4517	4638					
exon	5436	5559					
exon	6069	6251					
exon	6844	6960					
exon	7136	7923					
CDS	CHI3L1_cds.4	69 bp 3 exons	#109				
exon	1295	1319					
exon	1572	1601					
exon	2789	2802					
Allele	GB:Y08376_1	112 311 311	G>T				
	source	isSNP SNP00071934					
	consequence	CHI3L1_cds.1	106	Intron			
	consequence	CHI3L1_cds.2	107	Intron			
	consequence	CHI3L1_cds.3	108	Intron			
	consequence	CHI3L1_cds.4	109	3′			
Allele	GB:Y08376_1	112 438 438	A>G				
	source	isSNP SNP00008252					
	consequence	CHI3L1_cds.1	106	Missense	145-145	R>G	
	consequence	CHI3L1_cds.2	107	Missense	145-145	R>G	
	consequence	CHI3L1_cds.3	108	Missense	145-145	R>G	
	consequence	CHI3L1_cds.4	109	3′			
Allele	GB:Y08377_1	113 355 355	G>T				
	source	isSNP SNP00022932					
	consequence	CHI3L1_cds.1	106	Missense	174-174	L>I	
	consequence	CHI3L1_cds.2	107	Missense	174-174	L>I	
	consequence	CHI3L1_cds.3	108	Missense	174-174	L>I	
	consequence	CHI3L1_cds.4	109	3′			
Allele	GB:Y08378_1	114 506 506	A>G				
	source	isSNP SNP00005491					
	consequence	CHI3L1_cds.1	106	Intron			
	consequence	CHI3L1_cds.2	107	Intron			
	consequence	CHI3L1_cds.3	108	Intron			
	consequence	CHI3L1_cds.4	109	3′			
Allele	GB:Y08378_1	114 535 535	A>G				
	source	isSNP SNP00005492					
	consequence	CHI3L1_cds.1	106	Intron			
	consequence	CHI3L1_cds.2	107	Intron			
	consequence	CHI3L1_cds.3	108	Intron			
	consequence	CHI3L1_cds.4	109	3′			
Allele	GB:Y08378_1	114 641 641	A>G				
	source	isSNP SNP00028111					
	consequence	CHI3L1_cds.1	106	Intron			
	consequence	CHI3L1_cds.2	107	Intron			
	consequence	CHI3L1_cds.3	108	Intron			
	consequence	CHI3L1_cds.4	109	3′			
Allele	GB:Y08378_1	114 1560 1560	A>G				
	source	isSNP SNP00028112					
	consequence	CHI3L1_cds.1	106	Intron			
	consequence	CHI3L1_cds.2	107	Intron			

TABLE 1 (Cont.)

```
            consequence CHI3L1_cds.3        108     Intron
            consequence CHI3L1_cds.4        109     3'
  Allele    GB:Y08378_1 114    2163   2163  A>G
            source      isSNP  SNP00052666
            consequence CHI3L1_cds.1        106     3'
            consequence CHI3L1_cds.2        107     Silent       338-338      H
            consequence CHI3L1_cds.3        108     3'
            consequence CHI3L1_cds.4        109     3'
  Allele    GB:Y08378_1 114    2205   2205  A>G
            source      isSNP  SNP00072805
            consequence CHI3L1_cds.1        106     3'
            consequence CHI3L1_cds.2        107     Silent       352-352      A
            consequence CHI3L1_cds.3        108     3'
            consequence CHI3L1_cds.4        109     3'
  Allele    GB:Y08378_1 114    2602   2602  A>G
            source      isSNP  SNP00076686
            consequence CHI3L1_cds.1        106     3'
            consequence CHI3L1_cds.2        107     3'
            consequence CHI3L1_cds.3        108     3'
            consequence CHI3L1_cds.4        109     3'
  GIF CHI3L1-genomic-fwd.gif


CHI3L2
Full name : chitinase 3-like 2
Link : CHI3L2_link_cdna
  Subsequence      GB:HSU58514  1      1434    #117
  CDS GB:HSU58514.1       1173 bp              #118
      ORF      37     1209
  Allele    GB:HSU58514 117    412    412   A>G
            source      isSNP  SNP00021152
            consequence GB:HSU58514.1       118     Missense     126-126      N>D
  Allele    GB:HSU58514 117    581    581   A>G
            source      isSNP  SNP00021153
            consequence GB:HSU58514.1       118     Missense     182-182      A>V
  Allele    GB:HSU58514 117    972    972   A>G
            source      isSNP  SNP00115597
            consequence GB:HSU58514.1       118     Silent       312-312      K
  Allele    GB:HSU58514 117    1204   1204  A>G
            source      isSNP  SNP00068229
            consequence GB:HSU58514.1       118     Silent       390-390      L
  GIF CHI3L2-cdna-fwd.gif
Link : CHI3L2_alt_link_cdna
  Subsequence      GB:U58515_1  1      1500    #119
  CDS GB:U58515_1.1       1275 bp              #120
      ORF      1      1275
  Allele    GB:U58515_1 119    478    478   A>G
            source      isSNP  SNP00021152
            consequence GB:U58515_1.1       120     Missense     160-160      N>D
  Allele    GB:U58515_1 119    647    647   A>G
            source      isSNP  SNP00021153
            consequence GB:U58515_1.1       120     Missense     216-216      A>V
  Allele    GB:U58515_1 119    1038   1038  A>G
            source      isSNP  SNP00115597
```

TABLE 1 (Cont.)

```
            consequence  GB:U58515_1.1       120    Silent       346-346      K
  Allele    GB:U58515_1  119    1270  1270   A>G
            source       isSNP  SNP00068229
            consequence  GB:U58515_1.1       120    Silent       424-424      L
  GIF  CHI3L2-cdna-fwd.gif


CILP
Full name : cartilage intermediate layer protein
Link : CILP_link_cdna
  Subsequence       GB:AF035408  1      4175   #121
  CDS  GB:AF035408.1       3555 bp             #122
      ORF    130    3684
  Allele    GB:AF035408  121    430    430     A>G
            source       isSNP  SNP00123071
            consequence  GB:AF035408.1       122    Missense     101-101      P>S
  Allele    GB:AF035408  121    1677   1677    A>G
            source       isSNP  SNP00123072
            consequence  GB:AF035408.1       122    Silent       516-516      R
  Allele    GB:AF035408  121    3066   3066    A>G
            source       isSNP  SNP00020276
            consequence  GB:AF035408.1       122    Silent       979-979      R
  Allele    GB:AF035408  121    3263   3263    A>G
            source       isSNP  SNP00123073
            consequence  GB:AF035408.1       122    Missense     1045-1045    Y>C
  Allele    GB:AF035408  121    3625   3625    A>G
            source       isSNP  SNP00055164
            consequence  GB:AF035408.1       122    Missense     1166-1166    S>G
  GIF  CILP-cdna-fwd.gif
Link : CILP_link_genomic
  Subsequence       CILP_cds.1   3606   16639  #123
  Subsequence       GB:AB022430_1       1      19486  #124
  Subsequence       CILP_mrna_build.1  1911   17130  #125
  CDS  CILP_cds.1      3555 bp      8 exons      #123
      exon   3606   3666
      exon   5599   5691
      exon   6312   6581
      exon   7897   8076
      exon   8781   9095
      exon   9893   10001
      exon   11336  11493
      exon   14271  16639
  mRNA      CILP_mrna_build.1  4175 bp      9 exons      #125
      exon   1911   1933
      exon   3500   3666
      exon   5599   5691
      exon   6312   6581
      exon   7897   8076
      exon   8781   9095
      exon   9893   10001
      exon   11336  11493
      exon   14271  17130
  Allele    GB:AB022430_1      124    3567   3567   G>T
            source       wetSNP       GB:AB022430_1.v3567.A>C
```

TABLE 1 (Cont.)

```
            consequence  CILP_cds.1    123    5'
 Allele     GB:AB022430_1       124    6458   6458   A>G
            source       isSNP  SNP00123071
            consequence  CILP_cds.1    123    Missense     101-101     P>S
 Allele     GB:AB022430_1       124    9874   9874   A>G
            source       wetSNP        GB:AB022430_1.v9874.C>T
            consequence  CILP_cds.1    123    Intron
 Allele     GB:AB022430_1       124    9881   9881   A>G
            source       wetSNP        GB:AB022430_1.v9881.C>T
            consequence  CILP_cds.1    123    Intron
 Allele     GB:AB022430_1       124    11286  11286  A>T
            source       wetSNP        GB:AB022430_1.v11286.T>A
            consequence  CILP_cds.1    123    Intron
 Allele     GB:AB022430_1       124    11491  11491  A>G
            source       wetSNP        GB:AB022430_1.v11491.C>T
            consequence  CILP_cds.1    123    Missense     395-395     T>I
 Allele     GB:AB022430_1       124    14421  14421  C>G
            source       wetSNP        GB:AB022430_1.v14421.G>C
            consequence  CILP_cds.1    123    Missense     446-446     R>T
 Allele     GB:AB022430_1       124    14542  14542  A>G
            source       wetSNP        GB:AB022430_1.v14542.G>A
            consequence  CILP_cds.1    123    Silent       486-486     T
 Allele     GB:AB022430_1       124    14632  14632  A>G
            source       isSNP  SNP00123072
            consequence  CILP_cds.1    123    Silent       516-516     R
 Allele     GB:AB022430_1       124    15116  15116  A>G
            source       wetSNP        GB:AB022430_1.v15116.G>A
            consequence  CILP_cds.1    123    Missense     678-678     V>M
 Allele     GB:AB022430_1       124    15670  15670  A>G
            source       wetSNP        GB:AB022430_1.v15670.G>A
            consequence  CILP_cds.1    123    Silent       862-862     T
 Allele     GB:AB022430_1       124    16021  16021  A>G
            source       isSNP  SNP00020276
            consequence  CILP_cds.1    123    Silent       979-979     R
 Allele     GB:AB022430_1       124    16218  16218  A>G
            source       isSNP  SNP00123073
            consequence  CILP_cds.1    123    Missense     1045-1045   Y>C
 Allele     GB:AB022430_1       124    16580  16580  A>G
            source       isSNP  SNP00055164
            source       wetSNP        GB:AB022430_1.v16580.A>G
            consequence  CILP_cds.1    123    Missense     1166-1166   S>G
 GIF CILP-genomic-fwd.gif


COL10A1
Full name : collagen, type X, alpha 1
Link : COL10A1_link_cdna
  Subsequence      GB:X60382_1  1      3226    #126
  CDS GB:X60382_1.2        2043 bp             #127
      ORF     16     2058
  Allele      GB:X60382_1  126    95      95      A>G
              source       isSNP  SNP00034488
              consequence  GB:X60382_1.2         127    Missense     27-27  T>M
  Allele      GB:X60382_1  126    2294    2294    G>T
```

TABLE 1 (Cont.)

```
        source       isSNP SNP00113056
        consequence  GB:X60382_1.2       127    3'
    GIF COL10A1-cdna-fwd.gif



COL11A2
Full name : collagen, type XI, alpha 2
Link : FL_3421462_link_genomic
  Subsequence     GB:AL031228_1        1      175737       #128
  Subsequence     COL11A2_cds.1        93988  122550       #129
  Subsequence     COL11A2_cds.2        93988  122550       #130
  Subsequence     COL11A2_cds.3        93988  122550       #131
  Subsequence     COL11A2_cds.4        93988  122550       #132
  Subsequence     COL11A2_cds.5        93988  122550       #133
  Subsequence     COL11A2_cds.6        93988  122550       #134
  Subsequence     COL11A2_cds.7        93988  122550       #135
  Subsequence     COL11A2_cds.8        93988  122550       #136
  Subsequence     COL11A2_mrna_build.1      93988  122834       #137
  Subsequence     COL11A2_mrna_build.2      93988  122834       #138
  Subsequence     GB:AL031228_1.20     93762  123536       #139
  Subsequence     GB:AL031228_1.21     93988  122550       #140
  Subsequence     COL11A2_mrna_build.3      93769  125002       #141
  mRNA       GB:AL031228_1.20   6423 bp      66 exons     #139
      exon  93762  94069
      exon  96759  96908
      exon  97040  97250
      exon  97704  97866
      exon  99410  99601
      exon  100450      100527
      exon  101174      101236
      exon  101904      102083
      exon  105058      105117
      exon  105223      105264
      exon  105498      105560
      exon  105896      105970
      exon  106423      106509
      exon  106741      106797
      exon  106944      106997
      exon  107102      107155
      exon  107255      107308
      exon  107496      107549
      exon  107740      107793
      exon  107876      107920
      exon  108043      108096
      exon  108522      108566
      exon  108763      108816
      exon  109003      109047
      exon  109183      109236
      exon  109463      109507
      exon  109742      109795
      exon  109925      109969
      exon  110159      110212
      exon  110547      110654
      exon  111648      111701
```

TABLE 1 (Cont.)

	exon	112010	112063		
	exon	112173	112217		
	exon	112302	112355		
	exon	112483	112527		
	exon	112673	112726		
	exon	112827	112880		
	exon	113115	113168		
	exon	113591	113698		
	exon	113850	113939		
	exon	114125	114178		
	exon	114408	114515		
	exon	114654	114761		
	exon	114904	114957		
	exon	115061	115114		
	exon	115311	115418		
	exon	115618	115671		
	exon	115849	115902		
	exon	116128	116181		
	exon	116344	116397		
	exon	116738	116845		
	exon	117220	117273		
	exon	117469	117522		
	exon	117656	117709		
	exon	118376	118429		
	exon	118695	118802		
	exon	118911	118964		
	exon	119105	119158		
	exon	119401	119508		
	exon	119662	119715		
	exon	120022	120057		
	exon	120244	120297		
	exon	120412	120679		
	exon	121264	121376		
	exon	121755	121961		
	exon	122410	123536		
mRNA		COL11A2_mrna_build.3	6780 bp	66 exons	#141
	exon	93769	94341		
	exon	96759	96908		
	exon	97040	97250		
	exon	97704	97866		
	exon	99410	99601		
	exon	101174	101236		
	exon	101904	102083		
	exon	105058	105117		
	exon	105223	105264		
	exon	105498	105560		
	exon	105896	105970		
	exon	106423	106509		
	exon	106741	106797		
	exon	106944	106997		
	exon	107102	107155		
	exon	107255	107308		
	exon	107496	107549		
	exon	107740	107793		
	exon	107876	107920		

TABLE 1 (Cont.)

	exon	108043	108096		
	exon	108522	108566		
	exon	108763	108816		
	exon	109003	109047		
	exon	109183	109236		
	exon	109463	109507		
	exon	109742	109795		
	exon	109925	109969		
	exon	110159	110212		
	exon	110547	110654		
	exon	111648	111701		
	exon	112010	112063		
	exon	112173	112217		
	exon	112302	112355		
	exon	112483	112577		
	exon	112673	112726		
	exon	112827	112880		
	exon	113115	113168		
	exon	113591	113698		
	exon	113850	113939		
	exon	114125	114178		
	exon	114408	114515		
	exon	114654	114761		
	exon	114904	114957		
	exon	115061	115114		
	exon	115311	115418		
	exon	115618	115671		
	exon	115849	115902		
	exon	116128	116196		
	exon	116344	116397		
	exon	116738	116845		
	exon	117220	117273		
	exon	117469	117522		
	exon	117656	117709		
	exon	118376	118429		
	exon	118695	118802		
	exon	118911	118964		
	exon	119105	119158		
	exon	119401	119508		
	exon	120022	120057		
	exon	120244	120297		
	exon	120412	120679		
	exon	121264	121376		
	exon	121755	121961		
	exon	122183	122332		
	exon	122410	123530		
	exon	124988	125002		
CDS	COL11A2_cds.6		5157 bp	65 exons	#134
	exon	93988 94069			
	exon	96759 96908			
	exon	97040 97250			
	exon	97704 97866			
	exon	99410 99601			
	exon	100450	100527		
	exon	101174	101236		

TABLE 1 (Cont.)

exon	101904	102083
exon	105058	105117
exon	105223	105264
exon	105498	105560
exon	105896	105970
exon	106423	106509
exon	106741	106797
exon	106944	106997
exon	107102	107155
exon	107255	107308
exon	107496	107549
exon	107740	107793
exon	107876	107920
exon	108043	108096
exon	108522	108566
exon	108763	108816
exon	109003	109047
exon	109183	109236
exon	109463	109507
exon	109742	109795
exon	109925	109969
exon	110159	110212
exon	110547	110654
exon	111648	111701
exon	112010	112063
exon	112173	112217
exon	112302	112355
exon	112483	112527
exon	112673	112726
exon	112827	112880
exon	113115	113168
exon	113591	113698
exon	113850	113939
exon	114125	114178
exon	114408	114515
exon	114654	114761
exon	114904	114957
exon	115061	115114
exon	115311	115418
exon	115618	115671
exon	115849	115902
exon	116128	116181
exon	116344	116397
exon	116738	116845
exon	117220	117273
exon	117469	117522
exon	117656	117709
exon	118376	118429
exon	118695	118802
exon	118911	118964
exon	119105	119158
exon	119401	119508
exon	120022	120057
exon	120244	120297
exon	120412	120679

TABLE 1 (Cont.)

	exon	121264	121376		
	exon	121755	121961		
	exon	122410	122550		
CDS	GB:AL031228_1.21		5211 bp	66 exons	#140
	exon	93988	94069		
	exon	96759	96908		
	exon	97040	97250		
	exon	97704	97866		
	exon	99410	99601		
	exon	100450	100527		
	exon	101174	101236		
	exon	101904	102083		
	exon	105058	105117		
	exon	105223	105264		
	exon	105498	105560		
	exon	105896	105970		
	exon	106423	106509		
	exon	106741	106797		
	exon	106944	106997		
	exon	107102	107155		
	exon	107255	107308		
	exon	107496	107549		
	exon	107740	107793		
	exon	107876	107920		
	exon	108043	108096		
	exon	108522	108566		
	exon	108763	108816		
	exon	109003	109047		
	exon	109183	109236		
	exon	109463	109507		
	exon	109742	109795		
	exon	109925	109969		
	exon	110159	110212		
	exon	110547	110654		
	exon	111648	111701		
	exon	112010	112063		
	exon	112173	112217		
	exon	112302	112355		
	exon	112483	112527		
	exon	112673	112726		
	exon	112827	112880		
	exon	113115	113168		
	exon	113591	113698		
	exon	113850	113939		
	exon	114125	114178		
	exon	114408	114515		
	exon	114654	114761		
	exon	114904	114957		
	exon	115061	115114		
	exon	115311	115418		
	exon	115618	115671		
	exon	115849	115902		
	exon	116128	116181		
	exon	116344	116397		
	exon	116738	116845		

TABLE 1 (Cont.)

	exon	117220	117273		
	exon	117469	117522		
	exon	117656	117709		
	exon	118376	118429		
	exon	118695	118802		
	exon	118911	118964		
	exon	119105	119158		
	exon	119401	119508		
	exon	119662	119715		
	exon	120022	120057		
	exon	120244	120297		
	exon	120412	120679		
	exon	121264	121376		
	exon	121755	121961		
	exon	122410	122550		
CDS	COL11A2_cds.7		5049 bp	64 exons	#135
	exon	93988 94069			
	exon	96759 96908			
	exon	97040 97250			
	exon	97704 97866			
	exon	99410 99601			
	exon	101174	101236		
	exon	105058	105117		
	exon	105223	105264		
	exon	105498	105560		
	exon	105896	105970		
	exon	106423	106509		
	exon	106741	106797		
	exon	106944	106997		
	exon	107102	107155		
	exon	107255	107308		
	exon	107496	107549		
	exon	107740	107793		
	exon	107876	107920		
	exon	108043	108096		
	exon	108522	108566		
	exon	108763	108816		
	exon	109003	109047		
	exon	109183	109236		
	exon	109463	109507		
	exon	109742	109795		
	exon	109925	109969		
	exon	110159	110212		
	exon	110547	110654		
	exon	111648	111701		
	exon	112010	112063		
	exon	112173	112217		
	exon	112302	112355		
	exon	112483	112527		
	exon	112673	112726		
	exon	112827	112880		
	exon	113115	113168		
	exon	113591	113698		
	exon	113850	113939		
	exon	114125	114178		

TABLE 1 (Cont.)

	exon	114408	114515	
	exon	114654	114761	
	exon	114904	114957	
	exon	115061	115114	
	exon	115311	115418	
	exon	115618	115671	
	exon	115849	115902	
	exon	116128	116181	
	exon	116344	116397	
	exon	116738	116845	
	exon	117220	117273	
	exon	117469	117522	
	exon	117656	117709	
	exon	118376	118429	
	exon	118695	118802	
	exon	118911	118964	
	exon	119105	119158	
	exon	119401	119508	
	exon	120022	120057	
	exon	120244	120297	
	exon	120412	120679	
	exon	121264	121376	
	exon	121755	121961	
	exon	122183	122332	
	exon	122410	122550	
CDS	COL11A2_cds.8		4986 bp	63 exons #136
	exon	93988	94069	
	exon	96759	96908	
	exon	97040	97250	
	exon	97704	97866	
	exon	99410	99601	
	exon	105058	105117	
	exon	105223	105264	
	exon	105498	105560	
	exon	105896	105970	
	exon	106423	106509	
	exon	106741	106797	
	exon	106944	106997	
	exon	107102	107155	
	exon	107255	107308	
	exon	107496	107549	
	exon	107740	107793	
	exon	107876	107920	
	exon	108043	108096	
	exon	108522	108566	
	exon	108763	108816	
	exon	109003	109047	
	exon	109183	109236	
	exon	109463	109507	
	exon	109742	109795	
	exon	109925	109969	
	exon	110159	110212	
	exon	110547	110654	
	exon	111648	111701	
	exon	112010	112063	

TABLE 1 (Cont.)

	exon	112173	112217		
	exon	112302	112355		
	exon	112483	112527		
	exon	112673	112726		
	exon	112827	112880		
	exon	113115	113168		
	exon	113591	113698		
	exon	113850	113939		
	exon	114125	114178		
	exon	114408	114515		
	exon	114654	114761		
	exon	114904	114957		
	exon	115061	115114		
	exon	115311	115418		
	exon	115618	115671		
	exon	115849	115902		
	exon	116128	116181		
	exon	116344	116397		
	exon	116738	116845		
	exon	117220	117273		
	exon	117469	117522		
	exon	117656	117709		
	exon	118376	118429		
	exon	118695	118802		
	exon	118911	118964		
	exon	119105	119158		
	exon	119401	119508		
	exon	120022	120057		
	exon	120244	120297		
	exon	120412	120679		
	exon	121264	121376		
	exon	121755	121961		
	exon	122183	122332		
	exon	122410	122550		
CDS	COL11A2_cds.1		4890 bp	63 exons	#129
	exon	93988 94069			
	exon	96759 96908			
	exon	97040 97250			
	exon	97704 97866			
	exon	99410 99601			
	exon	105058	105117		
	exon	105223	105264		
	exon	105498	105560		
	exon	105896	105970		
	exon	106423	106509		
	exon	106741	106797		
	exon	106944	106997		
	exon	107102	107155		
	exon	107255	107308		
	exon	107496	107549		
	exon	107740	107793		
	exon	107876	107920		
	exon	108043	108096		
	exon	108522	108566		
	exon	108763	108816		

TABLE 1 (Cont.)

	exon	109003	109047		
	exon	109183	109236		
	exon	109463	109507		
	exon	109742	109795		
	exon	109925	109969		
	exon	110159	110212		
	exon	110547	110654		
	exon	111648	111701		
	exon	112010	112063		
	exon	112173	112217		
	exon	112302	112355		
	exon	112483	112527		
	exon	112673	112726		
	exon	112827	112880		
	exon	113115	113168		
	exon	113591	113698		
	exon	113850	113939		
	exon	114125	114178		
	exon	114408	114515		
	exon	114654	114761		
	exon	114904	114957		
	exon	115061	115114		
	exon	115311	115418		
	exon	115618	115671		
	exon	115849	115902		
	exon	116128	116181		
	exon	116344	116397		
	exon	116738	116845		
	exon	117220	117273		
	exon	117469	117522		
	exon	117656	117709		
	exon	118376	118429		
	exon	118695	118802		
	exon	118911	118964		
	exon	119105	119158		
	exon	119401	119508		
	exon	119662	119715		
	exon	120022	120057		
	exon	120244	120297		
	exon	120412	120679		
	exon	121264	121376		
	exon	121755	121961		
	exon	122410	122550		
CDS	COL11A2_cds.2		4953 bp	64 exons	#130
	exon	93988 94069			
	exon	96759 96908			
	exon	97040 97250			
	exon	97704 97866			
	exon	99410 99601			
	exon	101174	101236		
	exon	105058	105117		
	exon	105223	105264		
	exon	105498	105560		
	exon	105896	105970		
	exon	106423	106509		

TABLE 1 (Cont.)

	exon	106741	106797		
	exon	106944	106997		
	exon	107102	107155		
	exon	107255	107308		
	exon	107496	107549		
	exon	107740	107793		
	exon	107876	107920		
	exon	108043	108096		
	exon	108522	108566		
	exon	108763	108816		
	exon	109003	109047		
	exon	109183	109236		
	exon	109463	109507		
	exon	109742	109795		
	exon	109925	109969		
	exon	110159	110212		
	exon	110547	110654		
	exon	111648	111701		
	exon	112010	112063		
	exon	112173	112217		
	exon	112302	112355		
	exon	112483	112527		
	exon	112673	112726		
	exon	112827	112880		
	exon	113115	113168		
	exon	113591	113698		
	exon	113850	113939		
	exon	114125	114178		
	exon	114408	114515		
	exon	114654	114761		
	exon	114904	114957		
	exon	115061	115114		
	exon	115311	115418		
	exon	115618	115671		
	exon	115849	115902		
	exon	116128	116181		
	exon	116344	116397		
	exon	116738	116845		
	exon	117220	117273		
	exon	117469	117522		
	exon	117656	117709		
	exon	118376	118429		
	exon	118695	118802		
	exon	118911	118964		
	exon	119105	119158		
	exon	119401	119508		
	exon	119662	119715		
	exon	120022	120057		
	exon	120244	120297		
	exon	120412	120679		
	exon	121264	121376		
	exon	121755	121961		
	exon	122410	122550		
CDS	COL11A2_cds.3		5307 bp	66 exons	#131
	exon	93988 94069			

TABLE 1 (Cont.)

exon	96759	96908
exon	97040	97250
exon	97704	97866
exon	99410	99601
exon	100450	100527
exon	101174	101236
exon	101904	102083
exon	105058	105117
exon	105223	105264
exon	105498	105560
exon	105896	105970
exon	106423	106509
exon	106741	106797
exon	106944	106997
exon	107102	107155
exon	107255	107308
exon	107496	107549
exon	107740	107793
exon	107876	107920
exon	108043	108096
exon	108522	108566
exon	108763	108816
exon	109003	109047
exon	109183	109236
exon	109463	109507
exon	109742	109795
exon	109925	109969
exon	110159	110212
exon	110547	110654
exon	111648	111701
exon	112010	112063
exon	112173	112217
exon	112302	112355
exon	112483	112527
exon	112673	112726
exon	112827	112880
exon	113115	113168
exon	113591	113698
exon	113850	113939
exon	114125	114178
exon	114408	114515
exon	114654	114761
exon	114904	114957
exon	115061	115114
exon	115311	115418
exon	115618	115671
exon	115849	115902
exon	116128	116181
exon	116344	116397
exon	116738	116845
exon	117220	117273
exon	117469	117522
exon	117656	117709
exon	118376	118429
exon	118695	118802

TABLE 1 (Cont.)

	exon	118911	118964		
	exon	119105	119158		
	exon	119401	119508		
	exon	120022	120057		
	exon	120244	120297		
	exon	120412	120679		
	exon	121264	121376		
	exon	121755	121961		
	exon	122183	122332		
	exon	122410	122550		
mRNA		COL11A2_mrna_build.1	5174 bp	63 exons	#137
	exon	93988	94069		
	exon	96759	96908		
	exon	97040	97250		
	exon	97704	97866		
	exon	99410	99601		
	exon	105058	105117		
	exon	105223	105264		
	exon	105498	105560		
	exon	105896	105970		
	exon	106423	106509		
	exon	106741	106797		
	exon	106944	106997		
	exon	107102	107155		
	exon	107255	107308		
	exon	107496	107549		
	exon	107740	107793		
	exon	107876	107920		
	exon	108043	108096		
	exon	108522	108566		
	exon	108763	108816		
	exon	109003	109047		
	exon	109183	109236		
	exon	109463	109507		
	exon	109742	109795		
	exon	109925	109969		
	exon	110159	110212		
	exon	110547	110654		
	exon	111648	111701		
	exon	112010	112063		
	exon	112173	112217		
	exon	112302	112355		
	exon	112483	112527		
	exon	112673	112726		
	exon	112827	112880		
	exon	113115	113168		
	exon	113591	113698		
	exon	113850	113939		
	exon	114125	114178		
	exon	114408	114515		
	exon	114654	114761		
	exon	114904	114957		
	exon	115061	115114		
	exon	115311	115418		
	exon	115618	115671		

TABLE 1 (Cont.)

	exon	115849	115902		
	exon	116128	116181		
	exon	116344	116397		
	exon	116738	116845		
	exon	117220	117273		
	exon	117469	117522		
	exon	117656	117709		
	exon	118376	118429		
	exon	118695	118802		
	exon	118911	118964		
	exon	119105	119158		
	exon	119401	119508		
	exon	119662	119715		
	exon	120022	120057		
	exon	120244	120297		
	exon	120412	120679		
	exon	121264	121376		
	exon	121755	121961		
	exon	122410	122834		
CDS	COL11A2_cds.4		4836 bp	62 exons	#132
	exon	93988 94069			
	exon	96759 96908			
	exon	97040 97250			
	exon	97704 97866			
	exon	99410 99601			
	exon	105058	105117		
	exon	105223	105264		
	exon	105498	105560		
	exon	105896	105970		
	exon	106423	106509		
	exon	106741	106797		
	exon	106944	106997		
	exon	107102	107155		
	exon	107255	107308		
	exon	107496	107549		
	exon	107740	107793		
	exon	107876	107920		
	exon	108043	108096		
	exon	108522	108566		
	exon	108763	108816		
	exon	109003	109047		
	exon	109183	109236		
	exon	109463	109507		
	exon	109742	109795		
	exon	109925	109969		
	exon	110159	110212		
	exon	110547	110654		
	exon	111648	111701		
	exon	112010	112063		
	exon	112173	112217		
	exon	112302	112355		
	exon	112483	112527		
	exon	112673	112726		
	exon	112827	112880		
	exon	113115	113168		

TABLE 1 (Cont.)

exon	113591	113698			
exon	113850	113939			
exon	114125	114178			
exon	114408	114515			
exon	114654	114761			
exon	114904	114957			
exon	115061	115114			
exon	115311	115418			
exon	115618	115671			
exon	115849	115902			
exon	116128	116181			
exon	116344	116397			
exon	116738	116845			
exon	117220	117273			
exon	117469	117522			
exon	117656	117709			
exon	118376	118429			
exon	118695	118802			
exon	118911	118964			
exon	119105	119158			
exon	119401	119508			
exon	120022	120057			
exon	120244	120297			
exon	120412	120679			
exon	121264	121376			
exon	121755	121961			
exon	122410	122550			
mRNA	COL11A2_mrna_build.2		5237 bp	64 exons	#138
exon	93988	94069			
exon	96759	96908			
exon	97040	97250			
exon	97704	97866			
exon	99410	99601			
exon	101174	101236			
exon	105058	105117			
exon	105223	105264			
exon	105498	105560			
exon	105896	105970			
exon	106423	106509			
exon	106741	106797			
exon	106944	106997			
exon	107102	107155			
exon	107255	107308			
exon	107496	107549			
exon	107740	107793			
exon	107876	107920			
exon	108043	108096			
exon	108522	108566			
exon	108763	108816			
exon	109003	109047			
exon	109183	109236			
exon	109463	109507			
exon	109742	109795			
exon	109925	109969			
exon	110159	110212			

TABLE 1 (Cont.)

	exon	110547	110654	
	exon	111648	111701	
	exon	112010	112063	
	exon	112173	112217	
	exon	112302	112355	
	exon	112483	112527	
	exon	112673	112726	
	exon	112827	112880	
	exon	113115	113168	
	exon	113591	113698	
	exon	113850	113939	
	exon	114125	114178	
	exon	114408	114515	
	exon	114654	114761	
	exon	114904	114957	
	exon	115061	115114	
	exon	115311	115418	
	exon	115618	115671	
	exon	115849	115902	
	exon	116128	116181	
	exon	116344	116397	
	exon	116738	116845	
	exon	117220	117273	
	exon	117469	117522	
	exon	117656	117709	
	exon	118376	118429	
	exon	118695	118802	
	exon	118911	118964	
	exon	119105	119158	
	exon	119401	119508	
	exon	119662	119715	
	exon	120022	120057	
	exon	120244	120297	
	exon	120412	120679	
	exon	121264	121376	
	exon	121755	121961	
	exon	122410	122834	
CDS	COL11A2_cds.5	4899 bp	63 exons	#133
	exon	93988 94069		
	exon	96759 96908		
	exon	97040 97250		
	exon	97704 97866		
	exon	99410 99601		
	exon	101174	101236	
	exon	105058	105117	
	exon	105223	105264	
	exon	105498	105560	
	exon	105896	105970	
	exon	106423	106509	
	exon	106741	106797	
	exon	106944	106997	
	exon	107102	107155	
	exon	107255	107308	
	exon	107496	107549	
	exon	107740	107793	

TABLE 1 (Cont.)

exon	107876	107920			
exon	108043	108096			
exon	108522	108566			
exon	108763	108816			
exon	109003	109047			
exon	109183	109236			
exon	109463	109507			
exon	109742	109795			
exon	109925	109969			
exon	110159	110212			
exon	110547	110654			
exon	111648	111701			
exon	112010	112063			
exon	112173	112217			
exon	112302	112355			
exon	112483	112527			
exon	112673	112726			
exon	112827	112880			
exon	113115	113168			
exon	113591	113698			
exon	113850	113939			
exon	114125	114178			
exon	114408	114515			
exon	114654	114761			
exon	114904	114957			
exon	115061	115114			
exon	115311	115418			
exon	115618	115671			
exon	115849	115902			
exon	116128	116181			
exon	116344	116397			
exon	116738	116845			
exon	117220	117273			
exon	117469	117522			
exon	117656	117709			
exon	118376	118429			
exon	118695	118802			
exon	118911	118964			
exon	119105	119158			
exon	119401	119508			
exon	120022	120057			
exon	120244	120297			
exon	120412	120679			
exon	121264	121376			
exon	121755	121961			
exon	122410	122550			
Allele	GB:AL031228_1	128	122970	122970	A>G
	source	isSNP SNP00027609			
	consequence	COL11A2_cds.6	134	3′	
	consequence	GB:AL031228_1.21	140	3′	
	consequence	COL11A2_cds.7	135	3′	
	consequence	COL11A2_cds.8	136	3′	
	consequence	COL11A2_cds.1	129	3′	
	consequence	COL11A2_cds.2	130	3′	
	consequence	COL11A2_cds.3	131	3′	

TABLE 1 (Cont.)

```
            consequence  COL11A2_cds.4        132    3'
            consequence  COL11A2_cds.5        133    3'
  GIF  COL11A2-genomic-fwd.gif



COL9A2
Full name : collagen, type IX, alpha 2
Link : FL_3482334_link_cdna
  Subsequence        FN:3482334CB1       1       2864    #142
  CDS  FN:3482334CB1.1     2079 bp               #143
      ORF     99     2177
  Allele      FN:3482334CB1       142    1087    1087    A>G
              source        isSNP  SNP00032502
              consequence  FN:3482334CB1.1       143     Missense      330-330       Q>R
  Allele      FN:3482334CB1       142    1113    1113    C>G
              source        isSNP  SNP00107342
              consequence  FN:3482334CB1.1       143     Missense      339-339       L>V
  Allele      FN:3482334CB1       142    1301    1301    A>G
              source        isSNP  SNP00107343
              consequence  FN:3482334CB1.1       143     Silent        401-401       G
  Allele      FN:3482334CB1       142    1345    1345    C>G
              source        isSNP  SNP00107344
              consequence  FN:3482334CB1.1       143     Missense      416-416       G>A
  Allele      FN:3482334CB1       142    2211    2211    A>G
              source        isSNP  SNP00067542
              consequence  FN:3482334CB1.1       143     3'
  Allele      FN:3482334CB1       142    2317    2317    A>G
              source        isSNP  SNP00032503
              consequence  FN:3482334CB1.1       143     3'
  GIF  COL9A2-cdna-fwd.gif
 Link : FL_1651412_link_cdna
  Subsequence        FN:1651412CB1       1       2869    #144
  CDS  FN:1651412CB1.1     2067 bp               #145
      ORF     68     2134
  Allele      FN:1651412CB1       144    1044    1044    A>G
              source        isSNP  SNP00032502
              consequence  FN:1651412CB1.1       145     Missense      326-326       R>Q
  Allele      FN:1651412CB1       144    1070    1070    C>G
              source        isSNP  SNP00107342
              consequence  FN:1651412CB1.1       145     Missense      335-335       L>V
  Allele      FN:1651412CB1       144    1258    1258    A>G
              source        isSNP  SNP00107343
              consequence  FN:1651412CB1.1       145     Silent        397-397       G
  Allele      FN:1651412CB1       144    1302    1302    C>G
              source        isSNP  SNP00107344
              consequence  FN:1651412CB1.1       145     Missense      412-412       G>A
  Allele      FN:1651412CB1       144    2168    2168    A>G
              source        isSNP  SNP00067542
              consequence  FN:1651412CB1.1       145     3'
  Allele      FN:1651412CB1       144    2274    2274    A>G
              source        isSNP  SNP00032503
              consequence  FN:1651412CB1.1       145     3'
  GIF  COL9A2-cdna-fwd.gif
 Link : FL_1651412_link_genomic
```

TABLE 1 (Cont.)

```
Subsequence      GB:AF019406 1      17606 #146
Subsequence      GB:AF019406_1651412CD1   1115  17091 #147
Subsequence      GB:AF019406_3482334CD1   1115  17091 #148
Subsequence      FL_1651412_mrna_build.1  1048  17606 #149
Subsequence      FL_3482334_mrna_build.1  1017  17606 #150
mRNA      FL_1651412_mrna_build.1  2649 bp      32 exons      #149
     exon 1048  1189
     exon 2635  2709
     exon 3905  3940
     exon 4025  4087
     exon 5507  5560
     exon 5682  5717
     exon 5811  5834
     exon 6178  6231
     exon 6573  6626
     exon 6741  6788
     exon 7002  7058
     exon 7142  7195
     exon 7521  7574
     exon 7971  8024
     exon 8124  8177
     exon 8297  8350
     exon 10041 10094
     exon 10530 10583
     exon 10787 10840
     exon 12101 12145
     exon 12519 12572
     exon 13436 13489
     exon 13754 13807
     exon 13892 13963
     exon 14184 14219
     exon 14311 14355
     exon 14440 14472
     exon 14603 14749
     exon 15093 15147
     exon 15467 15655
     exon 16387 16464
     exon 16895 17606
CDS  GB:AF019406_3482334CD1   2079 bp      32 exons      #148
     exon 1115  1189
     exon 2635  2709
     exon 3905  3940
     exon 4025  4087
     exon 5507  5560
     exon 5682  5717
     exon 5811  5834
     exon 6178  6231
     exon 6573  6626
     exon 6741  6800
     exon 7002  7058
     exon 7142  7195
     exon 7521  7574
     exon 7971  8024
     exon 8124  8177
     exon 8297  8350
```

TABLE 1 (Cont.)

	exon	10041	10094		
	exon	10530	10583		
	exon	10787	10840		
	exon	12101	12145		
	exon	12519	12572		
	exon	13436	13489		
	exon	13754	13807		
	exon	13892	13963		
	exon	14184	14219		
	exon	14311	14355		
	exon	14440	14472		
	exon	14603	14749		
	exon	15093	15147		
	exon	15467	15655		
	exon	16387	16464		
	exon	16895	17091		
mRNA	FL_3482334_mrna_build.1	2692 bp		32 exons	#150
	exon	1017	1189		
	exon	2635	2709		
	exon	3905	3940		
	exon	4025	4087		
	exon	5507	5560		
	exon	5682	5717		
	exon	5811	5834		
	exon	6178	6231		
	exon	6573	6626		
	exon	6741	6800		
	exon	7002	7058		
	exon	7142	7195		
	exon	7521	7574		
	exon	7971	8024		
	exon	8124	8177		
	exon	8297	8350		
	exon	10041	10094		
	exon	10530	10583		
	exon	10787	10840		
	exon	12101	12145		
	exon	12519	12572		
	exon	13436	13489		
	exon	13754	13807		
	exon	13892	13963		
	exon	14184	14219		
	exon	14311	14355		
	exon	14440	14472		
	exon	14603	14749		
	exon	15093	15147		
	exon	15467	15655		
	exon	16387	16464		
	exon	16895	17606		
CDS	GB:AF019406_1651412CD1	2067 bp		32 exons	#147
	exon	1115	1189		
	exon	2635	2709		
	exon	3905	3940		
	exon	4025	4087		
	exon	5507	5560		

TABLE 1 (Cont.)

```
        exon    5682   5717
        exon    5811   5834
        exon    6178   6231
        exon    6573   6626
        exon    6741   6788
        exon    7002   7058
        exon    7142   7195
        exon    7521   7574
        exon    7971   8024
        exon    8124   8177
        exon    8297   8350
        exon    10041  10094
        exon    10530  10583
        exon    10787  10840
        exon    12101  12145
        exon    12519  12572
        exon    13436  13489
        exon    13754  13807
        exon    13892  13963
        exon    14184  14219
        exon    14311  14355
        exon    14440  14472
        exon    14603  14749
        exon    15093  15147
        exon    15467  15655
        exon    16387  16464
        exon    16895  17091
   Allele       GB:AF019406  146    10809  10809  A>G
                source       isSNP  SNP00032502
                consequence  GB:AF019406_3482334CD1    148    Missense     330-330
      Q>R
                consequence  GB:AF019406_1651412CD1    147    Missense     326-326
      Q>R
   Allele       GB:AF019406  146    13783  13783  A>G
                source       isSNP  SNP00107343
                consequence  GB:AF019406_3482334CD1    148    Silent       401-401
      G
                consequence  GB:AF019406_1651412CD1    147    Silent       397-397
      G
   Allele       GB:AF019406  146    17229  17229  A>G
                source       isSNP  SNP00032503
                consequence  GB:AF019406_3482334CD1    148    3'
                consequence  GB:AF019406_1651412CD1    147    3'
   GIF  COL9A2-genomic-fwd.gif


COMP
Full name : cartilage oligomeric matrix protein
Link : FL_1901242_link_cdna
   Subsequence       FN:1901242CB1       1       2447    #151
   CDS  FN:1901242CB1.1     2274 bp              #152
        ORF     23     2296
   Allele       FN:1901242CB1      151     1200    1200    A>G
                source       isSNP  SNP00017026
```

TABLE 1 (Cont.)

```
            consequence FN:1901242CB1.1     152    Missense     393-393     S>L
  Allele    FN:1901242CB1        151    1319   1319   C>G
            source        isSNP  SNP00108392
            consequence FN:1901242CB1.1     152    Missense     433-433     D>H
  Allele    FN:1901242CB1        151    1335   1335   C>G
            source        isSNP  SNP00017027
            consequence FN:1901242CB1.1     152    Missense     438-438     G>A
  Allele    FN:1901242CB1        151    1777   1777   A>G
            source        isSNP  SNP00017029
            consequence FN:1901242CB1.1     152    Silent       585-585     T
  GIF COMP-cdna-fwd.gif
Link : FL_1901242_link_genomic
  Subsequence       GB:AC003107  1      46275  #153
  Subsequence       GB:AC003107_1901242CD1     32077  23724  #154
  Subsequence       FL_1901242_mrna_build.1    32099  23582  #155
  CDS GB:AC003107_1901242CD1    2274 bp      19 exons     #154
       exon  32077  31999
       exon  31743  31658
       exon  31421  31370
       exon  30922  30750
       exon  30105  29968
       exon  29721  29647
       exon  29558  29400
       exon  29322  29218
       exon  29127  29020
       exon  28458  28299
       exon  27459  27341
       exon  27100  27048
       exon  26955  26774
       exon  26660  26482
       exon  26355  26307
       exon  25901  25705
       exon  25172  25000
       exon  24002  23863
       exon  23770  23724
  mRNA        FL_1901242_mrna_build.1  2438 bp      19 exons     #155
       exon  32099  31999
       exon  31743  31658
       exon  31421  31370
       exon  30922  30750
       exon  30105  29968
       exon  29721  29647
       exon  29558  29400
       exon  29322  29218
       exon  29127  29020
       exon  28458  28299
       exon  27459  27341
       exon  27100  27048
       exon  26955  26774
       exon  26660  26482
       exon  26355  26307
       exon  25901  25705
       exon  25172  25000
       exon  24002  23863
       exon  23770  23582
```

TABLE 1 (Cont.)

```
 Allele     GB:AC003107  153    25864  25864  A>G
            source       isSNP  SNP00017029
            consequence  GB:AC003107_1901242CD1   154    Silent        585-585
    T
 Allele     GB:AC003107  153    27417  27417  A>G
            source       isSNP  SNP00017026
            consequence  GB:AC003107_1901242CD1   154    Missense      393-393
    S>L
 Allele     GB:AC003107  153    32082  32082  A>G
            source       isSNP  SNP00017025
            consequence  GB:AC003107_1901242CD1   154    5'
 GIF  COMP-genomic-rev.gif


CRLF1
Full name : cytokine receptor-like factor 1
Link : CRLF1_link_cdna
  Subsequence      GB:AF073515_1      1      1804    #156
  CDS  GB:AF073515_1.1     1269  bp              #157
       ORF    204    1472
  Allele     GB:AF073515_1        156    984     984    A>G
             source        isSNP  SNP00015261
             consequence   GB:AF073515_1.1    157    Missense      261-261      P>S
  GIF  CRLF1-cdna-fwd.gif


CRP
Full name : C-reactive protein
Link : CRP_link_cdna
  Subsequence      GB:X56214_1  1      1631    #158
  CDS  GB:X56214_1.1       675  bp             #159
       ORF    90     764
  Allele     GB:X56214_1  158    447    447    A>G
             source       isSNP  SNP00100892
             consequence  GB:X56214_1.1        159    Missense      120-120      S>P
  Allele     GB:X56214_1  158    988    988    A>G
             source       isSNP  SNP00029575
             consequence  GB:X56214_1.1        159    3'
  Allele     GB:X56214_1  158    1010   1010   A>G
             source       isSNP  SNP00076237
             consequence  GB:X56214_1.1        159    3'
  Allele     GB:X56214_1  158    1146   1146   C>G
             source       isSNP  SNP00076238
             consequence  GB:X56214_1.1        159    3'
  Allele     GB:X56214_1  158    1175   1175   G>T
             source       isSNP  SNP00100893
             consequence  GB:X56214_1.1        159    3'
  Allele     GB:X56214_1  158    1406   1406   A>G
             source       isSNP  SNP00100894
             consequence  GB:X56214_1.1        159    3'
  Allele     GB:X56214_1  158    1525   1525   A>G
             source       isSNP  SNP00100895
             consequence  GB:X56214_1.1        159    3'
```

TABLE 1 (Cont.)

```
  GIF CRP-cdna-fwd.gif
Link : CRP_link_genomic
  Subsequence       GB:HUMCRPGA  1       2480   #160
  Allele      GB:HUMCRPGA  160    865     865    A>G
              source       isSNP  SNP00100892
  Allele      GB:HUMCRPGA  160    1404    1404   A>G
              source       isSNP  SNP00029575
  Allele      GB:HUMCRPGA  160    1426    1426   A>G
              source       isSNP  SNP00076237
  Allele      GB:HUMCRPGA  160    1562    1562   C>G
              source       isSNP  SNP00076238
  Allele      GB:HUMCRPGA  160    1591    1591   G>T
              source       isSNP  SNP00100893
  Allele      GB:HUMCRPGA  160    1822    1822   A>G
              source       isSNP  SNP00100894
  Allele      GB:HUMCRPGA  160    1941    1941   A>G
              source       isSNP  SNP00100895
  Allele      GB:HUMCRPGA  160    2045    2045   A>G
              source       isSNP  SNP00100896
  Allele      GB:HUMCRPGA  160    2159    2159   A>G
              source       isSNP  SNP00100897
  Allele      GB:HUMCRPGA  160    2260    2260   A>G
              source       isSNP  SNP00006286


CRTL1
Full name : cartilage linking protein 1
Link : CRTL1_link_cdna
  Subsequence       GB:HSU43328  1       1759   #161
  CDS  GB:HSU43328.1       1065 bp              #162
      ORF    118    1182
  Allele      GB:HSU43328  161    801     801    C>G
              source       isSNP  SNP00020236
              consequence  GB:HSU43328.1        162    Silent        228-228      G
  Allele      GB:HSU43328  161    1454    1454   A>G
              source       isSNP  SNP00002295
              consequence  GB:HSU43328.1        162    3'
  GIF CRTL1-cdna-fwd.gif


CTSC
Full name : cathepsin C
Link : CTSC_link_cdna
  Subsequence       GB:NM_001814         1       1838   #163
  CDS  GB:NM_001814.1      1392 bp              #164
      ORF    34     1425
  Allele      GB:NM_001814         163    491     491    A>G
              source       isSNP  SNP00006579
              consequence  GB:NM_001814.1       164    Missense      153-153      T>I
  Allele      GB:NM_001814         163    1206    1206   G>T
              source       isSNP  SNP00006580
              consequence  GB:NM_001814.1       164    Silent        391-391      T
  Allele      GB:NM_001814         163    1224    1224   A>G
```

TABLE 1 (Cont.)

```
            source        isSNP  SNP00105444
            consequence  GB:NM_001814.1     164    Silent       397-397      F
  GIF  CTSC-cdna-fwd.gif
Link : CTSC_link_genomic
  Subsequence        CTSC_cds.1   150285        106619        #165
  Subsequence        CTSC_cds.2   150285        106619        #166
  Subsequence        GB:AC011088_8      1       164991        #167
  Subsequence        CTSC_mrna_build.1  150318        106206        #168
  CDS  CTSC_cds.1   1392 bp      7  exons       #165
       exon  150285        150114
       exon  147695        147550
       exon  125167        125001
       exon  121931        121776
       exon  113258        113143
       exon  108877        108746
       exon  107121        106619
  CDS  CTSC_cds.2   1260 bp      6  exons       #166
       exon  150285        150114
       exon  147695        147550
       exon  125167        125001
       exon  121931        121776
       exon  113258        113143
       exon  107121        106619
  mRNA          CTSC_mrna_build.1  1838 bp      7  exons       #168
       exon  150318        150114
       exon  147695        147550
       exon  125167        125001
       exon  121931        121776
       exon  113258        113143
       exon  108877        108746
       exon  107121        106206
  Allele        GB:AC011088_8       167    106820        106820        A>G
                source        isSNP  SNP00105444
                consequence  CTSC_cds.1    165    Silent        397-397      F
                consequence  CTSC_cds.2    166    Silent        353-353      F
  Allele        GB:AC011088_8       167    106838        106838        G>T
                source        isSNP  SNP00006580
                consequence  CTSC_cds.1    165    Silent        391-391      T
                consequence  CTSC_cds.2    166    Silent        347-347      T
  Allele        GB:AC011088_8       167    122438        122438        A>G
                source        dbSNP  gnl|dbSNP|ss1078568_allele
                source        dbSNP  gnl|dbSNP|ss1088590_allele
                source        dbSNP  gnl|dbSNP|ss382670_allele
                source        dbSNP  gnl|dbSNP|ss403413_allele
                consequence  CTSC_cds.1    165    Intron
                consequence  CTSC_cds.2    166    Intron
  Allele        GB:AC011088_8       167    124932        124932        A>T
                source        wetSNP        GB:AC011088_8.v124932.A>T
                consequence  CTSC_cds.1    165    Intron
                consequence  CTSC_cds.2    166    Intron
  Allele        GB:AC011088_8       167    125028        125028        A>G
                source        isSNP  SNP00006579
                source        wetSNP        GB:AC011088_8.v125028.A>G
                consequence  CTSC_cds.1    165    Missense      153-153      I>T
                consequence  CTSC_cds.2    166    Missense      153-153      I>T
```

TABLE 1 (Cont.)

```
  Allele     GB:AC011088_8      167    142996      142996      A>G
             source       dbSNP  gnl|dbSNP|ss1530135_allele
             consequence  CTSC_cds.1   165    Intron
             consequence  CTSC_cds.2   166    Intron
  Allele     GB:AC011088_8      167    150261      150261      A>G
             source       wetSNP       GB:AC011088_8.v150261.G>A
             consequence  CTSC_cds.1   165    Missense     9-9   L>F
             consequence  CTSC_cds.2   166    Missense     9-9   L>F
  Allele     GB:AC011088_8      167    150303      150303      A>G
             source       isSNP  SNP00067426
             consequence  CTSC_cds.1   165    5'
             consequence  CTSC_cds.2   166    5'
  GIF  CTSC-genomic-rev.gif


CTSL
Full name : cathepsin L
Link : CTSL_link_genomic
  Subsequence      CTSL_cds.1   35962  179319       #169
  Subsequence      GB:AL160279_2       1      186528       #170
  Subsequence      CTSL_mrna_build.1  34477  179604       #171
  Subsequence      CTSL_cds.2   35962  179319       #172
  mRNA       CTSL_mrna_build.1  1577 bp      8 exons      #171
       exon   34477  34756
       exon   35952  36087
       exon   36385  36507
       exon   36608  36754
       exon   36943  37167
       exon   37931  38093
       exon   38739  38856
       exon   179220       179604
  CDS  CTSL_cds.1    1002 bp      7 exons      #169
       exon   35962  36087
       exon   36385  36507
       exon   36608  36754
       exon   36943  37167
       exon   37931  38093
       exon   38739  38856
       exon   179220       179319
  CDS  CTSL_cds.2    777 bp       6 exons      #172
       exon   35962  36087
       exon   36385  36507
       exon   36608  36754
       exon   37931  38093
       exon   38739  38856
       exon   179220       179319
  Allele     GB:AL160279_2      170    35919  35919  C>G
             source       wetSNP       GB:AL160279_2.v35919.C>G
             consequence  CTSL_cds.1   169    5'
             consequence  CTSL_cds.2   172    5'
  Allele     GB:AL160279_2      170    36118  36118  A>G
             source       wetSNP       GB:AL160279_2.v36118.C>T
             consequence  CTSL_cds.1   169    Intron
             consequence  CTSL_cds.2   172    Intron
```

TABLE 1 (Cont.)

```
Allele      GB:AL160279_2        170    36191  36191  G>T
            source        wetSNP        GB:AL160279_2.v36191.C>A
            consequence  CTSL_cds.1    169    Intron
            consequence  CTSL_cds.2    172    Intron
Allele      GB:AL160279_2        170    44998  44998  A>G
            source        isSNP  SNP00043782
            consequence  CTSL_cds.1    169    Intron
            consequence  CTSL_cds.2    172    Intron
Allele      GB:AL160279_2        170    45748  45748  A>G
            source        isSNP  SNP00007530
            consequence  CTSL_cds.1    169    Intron
            consequence  CTSL_cds.2    172    Intron
Allele      GB:AL160279_2        170    45833  45833  C>G
            source        isSNP  SNP00100366
            consequence  CTSL_cds.1    169    Intron
            consequence  CTSL_cds.2    172    Intron
Allele      GB:AL160279_2        170    46188  46188  A>G
            source        isSNP  SNP00100365
            consequence  CTSL_cds.1    169    Intron
            consequence  CTSL_cds.2    172    Intron
Allele      GB:AL160279_2        170    46599  46599  C>G
            source        isSNP  SNP00061067
            consequence  CTSL_cds.1    169    Intron
            consequence  CTSL_cds.2    172    Intron
Allele      GB:AL160279_2        170    46662  46662  C>G
            source        isSNP  SNP00100364
            consequence  CTSL_cds.1    169    Intron
            consequence  CTSL_cds.2    172    Intron
Allele      GB:AL160279_2        170    65760  65760  A>G
            source        isSNP  SNP00048929
            consequence  CTSL_cds.1    169    Intron
            consequence  CTSL_cds.2    172    Intron
Allele      GB:AL160279_2        170    81133  81133  A>G
            source        dbSNP  gnl|dbSNP|ss920176_allele
            source        dbSNP  gnl|dbSNP|ss1066694_allele
            source        dbSNP  gnl|dbSNP|ss402532_allele
            consequence  CTSL_cds.1    169    Intron
            consequence  CTSL_cds.2    172    Intron
Allele      GB:AL160279_2        170    104937        104937        A>G
            source        isSNP  SNP00055641
            consequence  CTSL_cds.1    169    Intron
            consequence  CTSL_cds.2    172    Intron
Allele      GB:AL160279_2        170    115466        115466        A>G
            source        isSNP  SNP00100363
            consequence  CTSL_cds.1    169    Intron
            consequence  CTSL_cds.2    172    Intron
Allele      GB:AL160279_2        170    127655        127655        A>T
            source        dbSNP  gnl|dbSNP|ss810769_allele
            consequence  CTSL_cds.1    169    Intron
            consequence  CTSL_cds.2    172    Intron
Allele      GB:AL160279_2        170    149731        149731        A>G
            source        dbSNP  gnl|dbSNP|ss1452230_allele
            consequence  CTSL_cds.1    169    Intron
            consequence  CTSL_cds.2    172    Intron
GIF  CTSL-genomic-fwd.gif
```

TABLE 1 (Cont.)

```
DAF
Full name : decay accelerating factor for complement
Link : DAF_link_genomic
  Subsequence      DAF_cds.1    131174      169024      #173
  Subsequence      DAF_cds.2    131174      169024      #174
  Subsequence      GB:AC031978_3      1     170170      #175
  Subsequence      DAF_mrna_build.1   131109      169897      #176
  CDS DAF_cds.1    1146 bp      10 exons     #173
      exon  131174       131273
      exon  131790       131975
      exon  133967       134158
      exon  135030       135129
      exon  136160       136245
      exon  140516       140704
      exon  146101       146226
      exon  146737       146817
      exon  148808       148828
      exon  168960       169024
  CDS DAF_cds.2    1125 bp      9 exons      #174
      exon  131174       131273
      exon  131790       131975
      exon  133967       134158
      exon  135030       135129
      exon  136160       136245
      exon  140516       140704
      exon  146101       146226
      exon  146737       146817
      exon  168960       169024
  mRNA        DAF_mrna_build.1   2084 bp      10 exons     #176
      exon  131109       131273
      exon  131790       131975
      exon  133967       134158
      exon  135030       135129
      exon  136160       136245
      exon  140516       140704
      exon  146101       146226
      exon  146737       146817
      exon  148808       148828
      exon  168960       169897
  Allele      GB:AC031978_3      175    132041      132041      A>G
              source       wetSNP      GB:AC031978_3.v132041.C>T
              consequence  DAF_cds.1    173    Intron
              consequence  DAF_cds.2    174    Intron
  Allele      GB:AC031978_3      175    146352      146352      A>G
              source       isSNP  SNP00072272
              consequence  DAF_cds.1    173    Intron
              consequence  DAF_cds.2    174    Intron
  Allele      GB:AC031978_3      175    146611      146611      A>G
              source       isSNP  SNP00072273
              consequence  DAF_cds.1    173    Intron
              consequence  DAF_cds.2    174    Intron
  Allele      GB:AC031978_3      175    146659      146659      A>G
              source       isSNP  SNP00030860
```

TABLE 1 (Cont.)

```
          consequence DAF_cds.1    173    Intron
          consequence DAF_cds.2    174    Intron
  Allele     GB:AC031978_3      175    165604        165604        A>G
          source       isSNP SNP00102533
          consequence DAF_cds.1    173    Intron
          consequence DAF_cds.2    174    Intron
  Allele     GB:AC031978_3      175    165743        165743        A>G
          source       isSNP SNP00102534
          consequence DAF_cds.1    173    Intron
          consequence DAF_cds.2    174    Intron
  GIF DAF-genomic-fwd.gif


E2F6
Full name : E2F transcription factor 6
Link : E2F6_link_cdna
  Subsequence      GB:AF041381 1      2027   #177
  Allele     GB:AF041381 177    1399   1399   A>G
          source       isSNP SNP00002319


EGF
Full name : EGF
Link : EGF_link_cdna
  Subsequence      GB:HSEGFRER 1      4871   #178
  CDS GB:HSEGFRER.1       3624 bp           #179
      ORF   437   4060
  Allele     GB:HSEGFRER 178    4453   4453   A>G
          source       isSNP SNP00043643
          consequence GB:HSEGFRER.1      179    3'
  GIF EGF-cdna-fwd.gif
Link : EGF_link_genomic
  Subsequence      GB:AC005509 1      143391        #180
  Subsequence      GB:AC004050 270590        143492        #181
  Subsequence      EGF_cds.1   64892 166730        #182
  Subsequence      EGF_mrna_build.1   64456 167538        #183
  CDS EGF_cds.1    3624 bp      24 exons     #182
      exon  64892 65018
      exon  92502 92701
      exon  94810 94991
      exon  95398 95625
      exon  96629 96831
      exon  110868       110993
      exon  112423       112545
      exon  113419       113541
      exon  114729       114854
      exon  115957       116093
      exon  120527       120675
      exon  126259       126363
      exon  127568       127791
      exon  131528       131695
      exon  132382       132531
      exon  134978       135097
```

TABLE 1 (Cont.)

exon	139300	139416					
exon	143859	143984					
exon	148522	148644					
exon	150008	150155					
exon	154954	155121					
exon	159780	159897					
exon	163427	163505					
exon	166477	166730					
mRNA	EGF_mrna_build.1	4868 bp		24 exons	#183		
exon	64456 65018						
exon	92502 92701						
exon	94810 94991						
exon	95398 95625						
exon	96629 96831						
exon	110868	110993					
exon	112423	112545					
exon	113419	113541					
exon	114729	114854					
exon	115957	116093					
exon	120527	120675					
exon	126259	126363					
exon	127568	127791					
exon	131528	131695					
exon	132382	132531					
exon	134978	135097					
exon	139300	139416					
exon	140140	140265					
exon	148522	148644					
exon	150008	150155					
exon	154954	155121					
exon	159780	159897					
exon	163427	163505					
exon	166477	167538					
Allele	GB:AC005509	180	70903	70903	A>G		
	source	dbSNP	gnl|dbSNP|ss875266_allele				
	consequence	EGF_cds.1		182	Intron		
Allele	GB:AC005509	180	92638	92638	A>G		
	source	wetSNP		GB:AC005509.v92638.C>T			
	consequence	EGF_cds.1		182	Silent	88-88	I
Allele	GB:AC005509	180	92670	92670	A>G		
	source	wetSNP		GB:AC005509.v92670.A>G			
	consequence	EGF_cds.1		182	Missense	99-99	Q>R
Allele	GB:AC005509	180	92763	92763	A>G		
	source	wetSNP		GB:AC005509.v92763.C>T			
	consequence	EGF_cds.1		182	Intron		
Allele	GB:AC005509	180	94933	94933	A>G		
	source	wetSNP		GB:AC005509.v94933.C>T			
	consequence	EGF_cds.1		182	Missense	151-151	H>Y
Allele	GB:AC005509	180	95444	95444	C>G		
	source	wetSNP		GB:AC005509.v95444.G>C			
	consequence	EGF_cds.1		182	Missense	186-186	D>H
Allele	GB:AC005509	180	96578	96578	G>T		
	source	wetSNP		GB:AC005509.v96578.A>C			
	consequence	EGF_cds.1		182	Intron		
Allele	GB:AC005509	180	96660	96660	C>G		

TABLE 1 (Cont.)

	source	wetSNP	GB:AC005509.v96660.G>C				
	consequence	EGF_cds.1	182	Missense	257-257	D>H	
Allele	GB:AC005509	180	96842	96842	A>G		
	source	wetSNP	GB:AC005509.v96842.G>A				
	consequence	EGF_cds.1	182	Intron			
Allele	GB:AC005509	180	96853	96853	A>G		
	source	wetSNP	GB:AC005509.v96853.G>A				
	consequence	EGF_cds.1	182	Intron			
Allele	GB:AC005509	180	100795	100795	G>T		
	source	dbSNP	gnl|dbSNP|ss48546_allele				
	source	dbSNP	gnl|dbSNP|ss569965_allele				
	consequence	EGF_cds.1	182	Intron			
Allele	GB:AC005509	180	112451	112451	A>G		
	source	wetSNP	GB:AC005509.v112451.T>C				
	consequence	EGF_cds.1	182	Silent	365-365	H	
Allele	GB:AC005509	180	113396	113396	A>G		
	source	wetSNP	GB:AC005509.v113396.T>C				
	consequence	EGF_cds.1	182	Intron			
Allele	GB:AC005509	180	113521	113521	A>G		
	source	wetSNP	GB:AC005509.v113521.G>A				
	consequence	EGF_cds.1	182	Missense	431-431	R>K	
Allele	GB:AC005509	180	114696	114696	A>G		
	source	wetSNP	GB:AC005509.v114696.C>T				
	consequence	EGF_cds.1	182	Intron			
Allele	GB:AC005509	180	126323	126323	A>G		
	source	wetSNP	GB:AC005509.v126323.A>G				
	consequence	EGF_cds.1	182	Missense	597-597	I>V	
Allele	GB:AC005509	180	127715	127715	A>G		
	source	wetSNP	GB:AC005509.v127715.C>T				
	consequence	EGF_cds.1	182	Silent	659-659	C	
Allele	GB:AC005509	180	131547	131547	A>G		
	source	wetSNP	GB:AC005509.v131547.A>G				
	consequence	EGF_cds.1	182	Silent	691-691	A	
Allele	GB:AC005509	180	131598	131598	A>G		
	source	wetSNP	GB:AC005509.v131598.G>A				
	consequence	EGF_cds.1	182	Missense	708-708	M>I	
Allele	GB:AC005509	180	131641	131641	C>G		
	source	wetSNP	GB:AC005509.v131641.G>C				
	consequence	EGF_cds.1	182	Missense	723-723	G>R	
Allele	GB:AC005509	180	132511	132511	A>T		
	source	wetSNP	GB:AC005509.v132511.A>T				
	consequence	EGF_cds.1	182	Missense	784-784	D>V	
Allele	GB:AC005509	180	139281	139281	A>G		
	source	wetSNP	GB:AC005509.v139281.G>A				
	consequence	EGF_cds.1	182	Intron			
Allele	GB:AC005509	180	139333	139333	A>G		
	source	wetSNP	GB:AC005509.v139333.T>C				
	consequence	EGF_cds.1	182	Missense	842-842	M>T	
Allele	GB:AC004050	181	126737	126737	G>T		
	source	wetSNP	GB:AC004050.v126737.C>A				
	consequence	EGF_cds.1	182	Intron			
Allele	GB:AC004050	181	122948	122948	A>G		
	source	isSNP	SNP00118827				
	consequence	EGF_cds.1	182	Intron			
Allele	GB:AC004050	181	122045	122045	A>T		

TABLE 1 (Cont.)

```
            source        wetSNP        GB:AC004050.v122045.A>T
            consequence   EGF_cds.1     182    Missense     920-920      E>V
  Allele    GB:AC004050   181    110980        110980       G>T
            source        isSNP  SNP00101773
            consequence   EGF_cds.1     182    Intron
  Allele    GB:AC004050   181    110796        110796       A>G
            source        wetSNP        GB:AC004050.v110796.A>G
            consequence   EGF_cds.1     182    Silent       1063-1063    L
  Allele    GB:AC004050   181    104082        104083       GC>GCC
            source        wetSNP        GB:AC004050.v104082.GC>GCC
            consequence   EGF_cds.1     182    Frameshift   1134-1135
  Allele    GB:AC004050   181    103468        103468       A>G
            source        isSNP  SNP00043643
            consequence   EGF_cds.1     182    3'
  GIF  EGF-genomic-fwd.gif


FDFT1
Full name : farnesyl-diphosphate farnesyltransferase 1
Link : FDFT1_link_cdna
  Subsequence       GB:FDFT1      1      1649   #184
  CDS  GB:FDFT1.1   1254 bp              #185
      ORF    45     1298
  Allele      GB:FDFT1      184    65     65     A>G
              source        isSNP  SNP00072434
              consequence   GB:FDFT1.1    185    Silent       7-7    L
  Allele      GB:FDFT1      184    178    178    A>G
              source        isSNP  SNP00065489
              consequence   GB:FDFT1.1    185    Missense     45-45  K>R
  Allele      GB:FDFT1      184    245    245    A>G
              source        isSNP  SNP00018570
              consequence   GB:FDFT1.1    185    Silent       67-67  N
  Allele      GB:FDFT1      184    590    590    A>G
              source        isSNP  SNP00123116
              consequence   GB:FDFT1.1    185    Silent       182-182      G
  Allele      GB:FDFT1      184    1016   1016   C>G
              source        isSNP  SNP00003188
              consequence   GB:FDFT1.1    185    Silent       324-324      L
  Allele      GB:FDFT1      184    1220   1220   A>G
              source        isSNP  SNP00123117
              consequence   GB:FDFT1.1    185    Silent       392-392      L
  Allele      GB:FDFT1      184    1532   1532   A>G
              source        isSNP  SNP00003189
              consequence   GB:FDFT1.1    185    3'
  GIF  FDFT1-cdna-fwd.gif
 Link : FDFT1_link_genomic
  Subsequence       FDFT1_cds.1  5681   37973  #186
  Subsequence       GB:AC025857_2_000033      1      19420  #187
  Subsequence       GB:AC025857_2_000021      19521  25487  #188
  Subsequence       GB:AC025857_2_000014      29099  25588  #189
  Subsequence       GB:AC025857_2_000029      29200  40859  #190
  Subsequence       FDFT1_mrna_build.1        5639   38324  #191
  mRNA        FDFT1_mrna_build.1        1647 bp       8 exons       #191
      exon    5639    5779
```

TABLE 1 (Cont.)

```
    exon   11642  11739
    exon   12515  12698
    exon   24238  24366
    exon   26209  26400
    exon   29608  29784
    exon   30882  31034
    exon   37752  38324
CDS FDFT1_cds.1  1254 bp      8 exons      #186
    exon   5681   5779
    exon   11642  11739
    exon   12515  12698
    exon   24238  24366
    exon   26209  26400
    exon   29608  29784
    exon   30882  31034
    exon   37752  37973
Allele      GB:AC025857_2_000033      187    5701   5701   A>G
            source        isSNP  SNP00072434
            consequence  FDFT1_cds.1  186    Silent        7-7    L
Allele      GB:AC025857_2_000033      187    6103   6103   C>G
            source        isSNP  SNP00072231
            consequence  FDFT1_cds.1  186    Intron
Allele      GB:AC025857_2_000033      187    11676  11676  A>G
            source        isSNP  SNP00065489
            consequence  FDFT1_cds.1  186    Missense      45-45  K>R
Allele      GB:AC025857_2_000014      189    2856   2856   A>G
            source        isSNP  SNP00123116
            consequence  FDFT1_cds.1  186    Silent        182-182     G
Allele      GB:AC025857_2_000029      190    1775   1775   C>G
            source        isSNP  SNP00003188
            consequence  FDFT1_cds.1  186    Silent        324-324     L
Allele      GB:AC025857_2_000029      190    5704   5704   A>G
            source        isSNP  SNP00096026
            consequence  FDFT1_cds.1  186    Intron
Allele      GB:AC025857_2_000029      190    8528   8528   A>G
            source        isSNP  SNP00105147
            consequence  FDFT1_cds.1  186    Intron
Allele      GB:AC025857_2_000029      190    8696   8696   A>G
            source        isSNP  SNP00123117
            consequence  FDFT1_cds.1  186    Silent        392-392     L
Allele      GB:AC025857_2_000029      190    9008   9008   A>G
            source        isSNP  SNP00003189
            consequence  FDFT1_cds.1  186    3'
Allele      GB:AC025857_2_000029      190    9148   9148   G>T
            source        isSNP  SNP00003190
            consequence  FDFT1_cds.1  186    3'
GIF FDFT1-genomic-fwd.gif


FGF1
Full name : Fibroblast growth factor 1 (acidic)
Link : FGF1_link_cdna
  Subsequence      GB:X51943_1  1      2259   #192
  CDS GB:X51943_1.1      468 bp               #193
```

TABLE 1 (Cont.)

```
    ORF     35      502
  Allele      GB:X51943_1  192    590    590    A>G
              source       isSNP  SNP00075582
              consequence  GB:X51943_1.1        193    3'
  Allele      GB:X51943_1  192    785    785    G>T
              source       isSNP  SNP00075583
              consequence  GB:X51943_1.1        193    3'
  Allele      GB:X51943_1  192    1855   1855   A>G
              source       isSNP  SNP00069845
              consequence  GB:X51943_1.1        193    3'
  Allele      GB:X51943_1  192    2007   2007   C>G
              source       isSNP  SNP00075584
              consequence  GB:X51943_1.1        193    3'
  GIF FGF1-cdna-fwd.gif
Link : FL_2535357_link_genomic
  Subsequence       GB:AC005370  1      76416  #194
  Subsequence       GB:AC005370_3284782CD1    45026  63860  #195
  Subsequence       FL_3284782_mrna_build.1   44979  67355  #196
  mRNA        FL_3284782_mrna_build.1  920 bp       4 exons       #196
       exon   44979  45194
       exon   58348  58451
       exon   63669  64259
       exon   67347  67355
  CDS  GB:AC005370_3284782CD1    465 bp       3 exons       #195
       exon   45026  45194
       exon   58348  58451
       exon   63669  63860
  Allele      GB:AC005370  194    63951  63951  A>G
              source       isSNP  SNP00075582
              consequence  GB:AC005370_3284782CD1    195    3'
  Allele      GB:AC005370  194    64146  64146  G>T
              source       isSNP  SNP00075583
              consequence  GB:AC005370_3284782CD1    195    3'
  Allele      GB:AC005370  194    65119  65119  G>T
              source       isSNP  SNP00012384
              consequence  GB:AC005370_3284782CD1    195    3'
  Allele      GB:AC005370  194    65217  65217  A>G
              source       isSNP  SNP00069845
              consequence  GB:AC005370_3284782CD1    195    3'
  Allele      GB:AC005370  194    65369  65369  C>G
              source       isSNP  SNP00075584
              consequence  GB:AC005370_3284782CD1    195    3'
  Allele      GB:AC005370  194    66005  66005  A>G
              source       isSNP  SNP00045433
              consequence  GB:AC005370_3284782CD1    195    3'
  GIF FGF1-genomic-fwd.gif


FGF2
Full name : fibroblast growth factor 2 (basic)
Link : FGF2_link_cdna
  Subsequence       GB:FGF2       1      6757   #197
  CDS GB:FGF2.1     633 bp              #198
      ORF    302    934
```

TABLE 1 (Cont.)

```
Allele      GB:FGF2       197    1651   1651   G>T
            source        isSNP  SNP00023270
            consequence   GB:FGF2.1     198    3'
Allele      GB:FGF2       197    1691   1691   A>G
            source        isSNP  SNP00058183
            consequence   GB:FGF2.1     198    3'
Allele      GB:FGF2       197    4603   4603   A>G
            source        isSNP  SNP00036340
            consequence   GB:FGF2.1     198    3'
Allele      GB:FGF2       197    4909   4909   A>G
            source        isSNP  SNP00036341
            consequence   GB:FGF2.1     198    3'
Allele      GB:FGF2       197    5455   5455   A>G
            source        isSNP  SNP00123025
            consequence   GB:FGF2.1     198    3'
Allele      GB:FGF2       197    5466   5466   C>G
            source        isSNP  SNP00036342
            consequence   GB:FGF2.1     198    3'
Allele      GB:FGF2       197    5892   5892   G>T
            source        isSNP  SNP00062439
            consequence   GB:FGF2.1     198    3'
Allele      GB:FGF2       197    5937   5937   A>G
            source        isSNP  SNP00062440
            consequence   GB:FGF2.1     198    3'
GIF  FGF2-cdna-fwd.gif


FGFR1
Full name : Fibroblast growth factor receptor-1
Link : FGFR1_link_cdna
  Subsequence        GB:M34185_1  1      3365    #199
  CDS  GB:M34185_1.1       2202 bp               #200
       ORF     256     2457
  Allele      GB:M34185_1  199    1471   1471    A>G
              source       isSNP  SNP00107960
              consequence  GB:M34185_1.1         200    Missense     406-406      A>T
  Allele      GB:M34185_1  199    3224   3224    G>T
              source       isSNP  SNP00107961
              consequence  GB:M34185_1.1         200    3'
  GIF  FGFR1-cdna-fwd.gif


FMOD
Full name : fibromodulin
Link : FMOD_link_cdna
  Subsequence        GB:FMOD       1      2863    #201
  CDS  GB:FMOD.1     1131 bp              #202
       ORF     21      1151
  Allele      GB:FMOD       201    2653   2653    C>G
              source        isSNP  SNP00001499
              consequence   GB:FMOD.1     202     3'
  Allele      GB:FMOD       201    2739   2739    A>G
              source        isSNP  SNP00001500
```

TABLE 1 (Cont.)

```
          consequence  GB:FMOD.1    202    3'
  GIF  FMOD-cdna-fwd.gif



FRZB
Full name : frizzled-related protein
Link : FRZB_link_cdna
  Subsequence       GB:U91903_1  1      1909   #203
  CDS  GB:U91903_1.1        978 bp             #204
       ORF    70     1047
  Allele       GB:U91903_1  203    667    667    A>G
               source       isSNP  SNP00016790
               consequence  GB:U91903_1.1        204    Missense     200-200       R>W
  Allele       GB:U91903_1  203    1039   1039   C>G
               source       isSNP  SNP00001065
               consequence  GB:U91903_1.1        204    Missense     324-324       R>G
  Allele       GB:U91903_1  203    1259   1259   A>G
               source       isSNP  SNP00001066
               consequence  GB:U91903_1.1        204    3'
  Allele       GB:U91903_1  203    1305   1305   A>G
               source       isSNP  SNP00016791
               consequence  GB:U91903_1.1        204    3'
  GIF  FRZB-cdna-fwd.gif



FST
Full name : Follistatin
Link : FST_link_cdna
  Subsequence       GB:FST        1      954    #205
  CDS  GB:FST.1      954 bp             #206
       ORF    1      954
  Allele       GB:FST        205    454    454    A>G
               source       isSNP  SNP00015508
               consequence  GB:FST.1      206    Missense     152-152       E>K
  Allele       GB:FST        205    853    853    C>G
               source       isSNP  SNP00052278
               consequence  GB:FST.1      206    Missense     285-285       A>P
  GIF  FST-cdna-fwd.gif
Link : FST_link_genomic
  Subsequence       FST_cds.1     77877  73442  #207
  Subsequence       GB:AC008901_2        1      192639        #208
  Subsequence       FST_mrna_build.1     77877  73440  #209
  CDS  FST_cds.1     951 bp        5 exons       #207
       exon   77877  77793
       exon   75788  75597
       exon   75164  74946
       exon   74599  74375
       exon   73671  73442
  mRNA         FST_mrna_build.1   953 bp        5 exons       #209
       exon   77877  77793
       exon   75788  75597
       exon   75164  74946
       exon   74599  74375
```

TABLE 1 (Cont.)

```
    exon   73671  73440
  Allele     GB:AC008901_2      208    73454  73454  A>G
             source        wetSNP      GB:AC008901_2.v73454.G>A
             consequence  FST_cds.1    207    Silent        313-313      S
  Allele     GB:AC008901_2      208    73540  73540  C>G
             source        isSNP  SNP00052278
             consequence  FST_cds.1    207    Missense      285-285      A>P
  Allele     GB:AC008901_2      208    74988  74988  A>G
             source        isSNP  SNP00015508
             consequence  FST_cds.1    207    Missense      152-152      E>K
  Allele     GB:AC008901_2      208    76361  76361  C>G
             source        dbSNP  gnl|dbSNP|ss42460_allele
             consequence  FST_cds.1    207    Intron
  Allele     GB:AC008901_2      208    76373  76373  A>G
             source        dbSNP  gnl|dbSNP|ss1048607_allele
             source        dbSNP  gnl|dbSNP|ss226044_allele
             consequence  FST_cds.1    207    Intron
  Allele     GB:AC008901_2      208    76384  76384  A>G
             source        dbSNP  gnl|dbSNP|ss839844_allele
             consequence  FST_cds.1    207    Intron
  GIF  FST-genomic-rev.gif


G0S2
Full name : putative lymphocyte G0\/G1 switch gene
Link : FL_3732868_link_genomic
  Subsequence      GB:HS28010   1       97700  #210
  Subsequence      GB:HS28010_3732868CD1      52369  52680  #211
  Subsequence      FL_3732868_mrna_build.1   52008  53073  #212
  mRNA       FL_3732868_mrna_build.1   963 bp       2 exons      #212
      exon   52008  52233
      exon   52337  53073
  CDS  GB:HS28010_3732868CD1     312  bp       1  exon       #211
      exon   52369  52680
  Allele     GB:HS28010   210    52341  52341  A>G
             source        isSNP  SNP00039143
             source        wetSNP      GB:HS28010.v52341.T>C
             consequence  GB:HS28010_3732868CD1     211    5'
  GIF  G0S2-genomic-fwd.gif


GADD34
Full name : growth arrest and DNA-damage-inducible 34
Link : GADD34_link_cdna
  Subsequence      GB:HSU83981  1       2331   #213
  CDS  GB:HSU83981.1       2025 bp              #214
      ORF    223    2247
  Allele     GB:HSU83981  213    205    205    A>G
             source        isSNP  SNP00116263
             consequence  GB:HSU83981.1        214    5'
  Allele     GB:HSU83981  213    314    314    A>G
             source        isSNP  SNP00116264
             consequence  GB:HSU83981.1        214    Missense      31-31  R>H
```

TABLE 1 (Cont.)

Allele	GB:HSU83981	213	316	316	A>G		
	source	isSNP	SNP00029694				
	consequence	GB:HSU83981.1	214	Missense	32-32	A>T	
Allele	GB:HSU83981	213	974	974	C>G		
	source	isSNP	SNP00006368				
	consequence	GB:HSU83981.1	214	Missense	251-251	R>P	
Allele	GB:HSU83981	213	1051	1051	A>G		
	source	isSNP	SNP00006369				
	consequence	GB:HSU83981.1	214	Missense	277-277	K>E	
Allele	GB:HSU83981	213	1156	1156	A>G		
	source	isSNP	SNP00006370				
	consequence	GB:HSU83981.1	214	Missense	312-312	G>S	
Allele	GB:HSU83981	213	1605	1605	A>G		
	source	isSNP	SNP00069978				
	consequence	GB:HSU83981.1	214	Silent	461-461	L	
Allele	GB:HSU83981	213	1650	1650	G>T		
	source	isSNP	SNP00069979				
	consequence	GB:HSU83981.1	214	Missense	476-476	R>S	
Allele	GB:HSU83981	213	2011	2011	A>G		
	source	isSNP	SNP00006372				
	consequence	GB:HSU83981.1	214	Missense	597-597	T>A	
Allele	GB:HSU83981	213	2184	2184	A>G		
	source	isSNP	SNP00006373				
	consequence	GB:HSU83981.1	214	Silent	654-654	A	
Allele	GB:HSU83981	213	2199	2199	C>G		
	source	isSNP	SNP00006374				
	consequence	GB:HSU83981.1	214	Silent	659-659	S	
GIF GADD34-cdna-fwd.gif							
Link : GADD34_link_genomic							
Subsequence	GADD34_cds.1	221390	224129	#215			
Subsequence	GB:AC026803_2	1	247509	#216			
Subsequence	GADD34_mrna_build.1	220595	224213	#217			
mRNA	GADD34_mrna_build.1	2331 bp	3 exons	#217			
exon	220595	220807					
exon	221381	223054					
exon	223770	224213					
CDS	GADD34_cds.1	2025 bp	2 exons	#215			
exon	221390	223054					
exon	223770	224129					
Allele	GB:AC026803_2	216	221481	221481	A>G		
	source	isSNP	SNP00116264				
	consequence	GADD34_cds.1	215	Missense	31-31	R>H	
Allele	GB:AC026803_2	216	221483	221483	A>G		
	source	isSNP	SNP00029694				
	consequence	GADD34_cds.1	215	Missense	32-32	A>T	
Allele	GB:AC026803_2	216	221941	221941	A>G		
	source	wetSNP	GB:AC026803_2.v221941.G>A				
	consequence	GADD34_cds.1	215	Silent	184-184	R	
Allele	GB:AC026803_2	216	221985	221985	A>G		
	source	wetSNP	GB:AC026803_2.v221985.T>C				
	consequence	GADD34_cds.1	215	Missense	199-199	V>A	
Allele	GB:AC026803_2	216	222141	222141	C>G		
	source	isSNP	SNP00006368				
	source	wetSNP	GB:AC026803_2.v222141.G>C				
	consequence	GADD34_cds.1	215	Missense	251-251	R>P	

TABLE 1 (Cont.)

```
  Allele      GB:AC026803_2       216    222218        222218        A>G
              source        isSNP SNP00006369
              consequence  GADD34_cds.1         215    Missense      277-277       K>E
  Allele      GB:AC026803_2       216    222323        222323        A>G
              source        isSNP SNP00006370
              consequence  GADD34_cds.1         215    Missense      312-312       G>S
  Allele      GB:AC026803_2       216    222772        222772        A>G
              source        isSNP SNP00069978
              consequence  GADD34_cds.1         215    Silent        461-461       L
  Allele      GB:AC026803_2       216    222817        222817        G>T
              source        isSNP SNP00069979
              consequence  GADD34_cds.1         215    Missense      476-476       R>S
  Allele      GB:AC026803_2       216    223893        223893        A>G
              source        isSNP SNP00006372
              consequence  GADD34_cds.1         215    Missense      597-597       T>A
  Allele      GB:AC026803_2       216    224066        224066        A>G
              source        isSNP SNP00006373
              consequence  GADD34_cds.1         215    Silent        654-654       A
  Allele      GB:AC026803_2       216    224081        224081        C>G
              source        isSNP SNP00006374
              consequence  GADD34_cds.1         215    Silent        659-659       S
  GIF GADD34-genomic-fwd.gif


GLI
Full name : glioma-associated oncogene homolog
Link : GLI_link_cdna
  Subsequence      GB:NM_005269_1      1      3600    #218
  CDS GB:NM_005269_1.1    3321 bp             #219
      ORF     79     3399
  Allele      GB:NM_005269_1      218    2179   2179   A>G
              source        isSNP SNP00018615
              consequence  GB:NM_005269_1.1    219    Missense      701-701       R>G
  Allele      GB:NM_005269_1      218    2202   2202   A>G
              source        isSNP SNP00072776
              consequence  GB:NM_005269_1.1    219    Silent        708-708       E
  Allele      GB:NM_005269_1      218    2876   2876   A>G
              source        isSNP SNP00112595
              consequence  GB:NM_005269_1.1    219    Missense      933-933       G>D
  Allele      GB:NM_005269_1      218    3243   3243   C>G
              source        isSNP SNP00018616
              consequence  GB:NM_005269_1.1    219    Missense      1055-1055     E>D
  Allele      GB:NM_005269_1      218    3376   3376   C>G
              source        isSNP SNP00018617
              consequence  GB:NM_005269_1.1    219    Missense      1100-1100     E>Q
  GIF GLI-cdna-fwd.gif


GLI3
Full name : GLI-Kruppel family member GLI3
Link : GLI3_link_cdna
  Subsequence      GB:NM_000168_1      1      5046    #220
  CDS GB:NM_000168_1.1    4791 bp             #221
```

TABLE 1 (Cont.)

```
      ORF    55     4845
  Allele     GB:NM_000168_1     220    4502   4502   A>G
             source       isSNP  SNP00031650
             consequence  GB:NM_000168_1.1    221    Missense    1483-1483    G>D
  Allele     GB:NM_000168_1     220    4663   4663   A>G
             source       isSNP  SNP00073523
             consequence  GB:NM_000168_1.1    221    Missense    1537-1537    R>C
  GIF  GLI3-cdna-fwd.gif


HAS1
Full name : hyaluronan synthase 1
Link : HAS1_link_cdna
  Subsequence      GB:NM_001523       1      2088   #222
  CDS GB:NM_001523.1      1737 bp            #223
      ORF    36     1772
  Allele     GB:NM_001523       222    75     75     A>G
             source       isSNP  SNP00096015
             consequence  GB:NM_001523.1      223    Missense    14-14  R>C
  Allele     GB:NM_001523       222    1889   1889   G>T
             source       isSNP  SNP00064738
             consequence  GB:NM_001523.1      223    3'
  GIF  HAS1-cdna-fwd.gif
Link : HAS1_link_genomic
  Subsequence      HAS1_cds.1   153154        142648       #224
  Subsequence      GB:AC018755_2      1       231222       #225
  Subsequence      HAS1_mrna_build.1  153189        142333       #226
  CDS HAS1_cds.1    1737 bp       5 exons      #224
      exon   153154       153146
      exon   149119       148427
      exon   146414       146189
      exon   145609       145477
      exon   143323       142648
  mRNA       HAS1_mrna_build.1  2087 bp      5 exons      #226
      exon   153189       153146
      exon   149119       148427
      exon   146414       146189
      exon   145609       145477
      exon   143323       142333
  Allele     GB:AC018755_2      225    142531        142531       G>T
             source       isSNP  SNP00064738
             consequence  HAS1_cds.1    224    3'
  Allele     GB:AC018755_2      225    147775        147775       G>T
             source       dbSNP  gnl|dbSNP|ss715930_allele
             consequence  HAS1_cds.1    224    Intron
  Allele     GB:AC018755_2      225    149089        149089       A>G
             source       isSNP  SNP00096015
             consequence  HAS1_cds.1    224    Missense    14-14  C>R
  Allele     GB:AC018755_2      225    149293        149293       C>G
             source       dbSNP  gnl|dbSNP|ss713606_allele
             consequence  HAS1_cds.1    224    Intron
  GIF  HAS1-genomic-rev.gif
```

TABLE 1 (Cont.)

HAS2
Full name : hyaluronan synthase 2
Link : HAS2_link_cdna

Subsequence	GB:NM_005328	1		3003	#227		
CDS GB:NM_005328.1	1659 bp			#228			
ORF	536	2194					
Allele	GB:NM_005328	227	381	381	A>G		
	source	isSNP	SNP00072998				
	consequence	GB:NM_005328.1		228	5'		
Allele	GB:NM_005328	227	1357	1357	G>T		
	source	isSNP	SNP00104961				
	consequence	GB:NM_005328.1		228	Missense	274-274	F>L
GIF HAS2-cdna-fwd.gif							

HSPG2
Full name : heparan sulfate proteoglycan 2
Link : HSPG2_link_cdna

Subsequence	GB:NM_005529_2	1		13793	#229		
CDS GB:NM_005529_2.1	13182 bp			#230			
ORF	41	13222					
Allele	GB:NM_005529_2	229	2155	2155	A>G		
	source	isSNP	SNP00054627				
	consequence	GB:NM_005529_2.1		230	Silent	705-705	A
Allele	GB:NM_005529_2	229	2340	2340	A>G		
	source	isSNP	SNP00054628				
	consequence	GB:NM_005529_2.1		230	Missense	767-767	S>N
Allele	GB:NM_005529_2	229	3603	3603	A>G		
	source	isSNP	SNP00109135				
	consequence	GB:NM_005529_2.1		230	Missense	1188-1188	R>Q
Allele	GB:NM_005529_2	229	3734	3734	A>G		
	source	isSNP	SNP00109136				
	consequence	GB:NM_005529_2.1		230	Missense	1232-1232	G>S
Allele	GB:NM_005529_2	229	3943	3943	A>G		
	source	isSNP	SNP00054629				
	consequence	GB:NM_005529_2.1		230	Silent	1301-1301	V
Allele	GB:NM_005529_2	229	4032	4032	A>G		
	source	isSNP	SNP00054630				
	consequence	GB:NM_005529_2.1		230	Missense	1331-1331	G>D
Allele	GB:NM_005529_2	229	4554	4554	A>G		
	source	isSNP	SNP00109138				
	consequence	GB:NM_005529_2.1		230	Missense	1505-1505	V>A
Allele	GB:NM_005529_2	229	7042	7042	A>G		
	source	isSNP	SNP00048871				
	consequence	GB:NM_005529_2.1		230	Silent	2334-2334	N
Allele	GB:NM_005529_2	229	7503	7503	A>G		
	source	isSNP	SNP00109139				
	consequence	GB:NM_005529_2.1		230	Missense	2488-2488	S>L
Allele	GB:NM_005529_2	229	9548	9548	A>G		
	source	isSNP	SNP00109140				
	consequence	GB:NM_005529_2.1		230	Missense	3170-3170	T>A
Allele	GB:NM_005529_2	229	10294	10294	A>G		
	source	isSNP	SNP00109141				
	consequence	GB:NM_005529_2.1		230	Silent	3418-3418	S

TABLE 1 (Cont.)

```
  Allele      GB:NM_005529_2      229    10663  10663  A>G
              source        isSNP  SNP00109142
              consequence  GB:NM_005529_2.1    230    Silent       3541-3541    V
  Allele      GB:NM_005529_2      229    10941  10941  A>G
              source        isSNP  SNP00109143
              consequence  GB:NM_005529_2.1    230    Missense     3634-3634    Q>R
  Allele      GB:NM_005529_2      229    11233  11233  G>T
              source        isSNP  SNP00009830
              consequence  GB:NM_005529_2.1    230    Silent       3731-3731    V
  Allele      GB:NM_005529_2      229    12358  12358  A>G
              source        isSNP  SNP00009831
              consequence  GB:NM_005529_2.1    230    Silent       4106-4106    D
  Allele      GB:NM_005529_2      229    12604  12604  A>G
              source        isSNP  SNP00038416
              consequence  GB:NM_005529_2.1    230    Silent       4188-4188    S
  GIF HSPG2-cdna-fwd.gif


IBSP
Full name : IBSP
Link : IBSP_link_cdna
  Subsequence       GB:HUMSIALO  1      1037    #231
  CDS GB:HUMSIALO.1        954 bp               #232
      ORF    72     1025
  Allele      GB:HUMSIALO  231    494    494    A>G
              source        isSNP  SNP00065793
              consequence  GB:HUMSIALO.1       232    Silent       141-141      N
  Allele      GB:HUMSIALO  231    655    655    A>G
              source        isSNP  SNP00065794
              consequence  GB:HUMSIALO.1       232    Missense     195-195      G>E
  Allele      GB:HUMSIALO  231    709    709    A>G
              source        isSNP  SNP00018906
              consequence  GB:HUMSIALO.1       232    Missense     213-213      G>D
  GIF IBSP-cdna-fwd.gif
Link : IBSP_link_genomic
  Subsequence       GB:HUMBNSP01        1      2415   #233
  Subsequence       GB:HUMBNSP02        2516   3359   #234
  Subsequence       GB:HUMBNSP03        3460   5094   #235
  Subsequence       GB:HUMBNSP04        5195   9497   #236
  Subsequence       IBSP_cds.1   2863   7195   #237
  CDS IBSP_cds.1    954 bp       6 exons       #237
      exon   2863   2916
      exon   3009   3059
      exon   3158   3235
      exon   3571   3633
      exon   5882   6040
      exon   6647   7195
  Allele      GB:HUMBNSP04        236    1631   1631   A>G
              source        isSNP  SNP00065794
              consequence  IBSP_cds.1    237    Missense     195-195      E>G
  Allele      GB:HUMBNSP04        236    1685   1685   A>G
              source        isSNP  SNP00018906
              consequence  IBSP_cds.1    237    Missense     213-213      G>D
  GIF IBSP-genomic-fwd.gif
```

TABLE 1 (Cont.)

```
IER3
Full name : immediate early response 3
Link : IER3_link_cdna
  Subsequence      GB:Y14551_1  1      1230   #238
  CDS  GB:Y14551_1.1      471 bp              #239
      ORF    12     482
  Allele      GB:Y14551_1  238    838    838    A>G
              source       isSNP  SNP00052893
              consequence  GB:Y14551_1.1        239    3'
  GIF  IER3-cdna-fwd.gif
Link : FL_758754_link_genomic
  Subsequence      GB:AC006165  1      44118  #240
  Subsequence      GB:AC006165_2619577CD1    14601  15183  #241
  Subsequence      FL_2619577_mrna_build.1   14585  15920  #242
  mRNA        FL_2619577_mrna_build.1  1224 bp      2 exons      #242
      exon   14585  14810
      exon   14923  15920
  CDS  GB:AC006165_2619577CD1    471 bp       2 exons      #241
      exon   14601  14810
      exon   14923  15183
  Allele      GB:AC006165  240    15539  15539  A>G
              source       isSNP  SNP00052893
              consequence  GB:AC006165_2619577CD1    241    3'
  GIF  IER3-genomic-fwd.gif


IHH
Full name : IHH
Link : IHH_link_cdna
  Subsequence      GB:HUMIHH    1      1277   #243
  CDS  GB:HUMIHH.2  939 bp              #244
      ORF    2      940
  Allele      GB:HUMIHH    243    457    457    A>G
              source       isSNP  SNP00097225
              consequence  GB:HUMIHH.2  244    Silent       152-152      P
  GIF  IHH-cdna-fwd.gif
Link : IHH_link_genomic
  Subsequence      IHH_cds.1    1      1469   #245
  Subsequence      GB:AB010092_1       1      315    #246
  Subsequence      GB:AB018075_1       416    698    #247
  Subsequence      GB:AB018076_1       799    1481   #248
  CDS  IHH_cds.1    1236 bp      3 exons      #245
      exon   1      315
      exon   426    687
      exon   811    1469
  Allele      GB:AB018075_1      247    194    194    A>G
              source       wetSNP        GB:AB018075_1.v194.G>A
              consequence  IHH_cds.1    245    Missense     167-167      A>T
  Allele      GB:AB018076_1      248    188    188    A>G
              source       isSNP  SNP00097225
              consequence  IHH_cds.1    245    Silent       251-251      P
  GIF  IHH-genomic-fwd.gif
```

TABLE 1 (Cont.)

```
INHBA
Full name : inhibin, beta A
Link : FL_3526170_link_cdna
  Subsequence      FN:3526170CB1      1      1620   #249
  CDS  FN:3526170CB1.1     1281 bp           #250
      ORF    216    1496
  Allele      FN:3526170CB1       249    607    607    G>T
              source        isSNP  SNP00068777
              consequence  FN:3526170CB1.1     250    Missense     131-131      T>K
  GIF  INHBA-cdna-fwd.gif
Link : FL_3526170_link_genomic
  Subsequence      GB:AC005027  1       199878       #251
  Subsequence      GB:AC005027_3526170CD1    16865  54957  #252
  Subsequence      FL_3526170_mrna_build.1   14163  55081  #253
  mRNA        FL_3526170_mrna_build.1   1620 bp      3  exons      #253
      exon   14163  14234
      exon   16722  17252
      exon   54065  55081
  CDS  GB:AC005027_3526170CD1    1281 bp       2  exons      #252
      exon   16865  17252
      exon   54065  54957
  Allele      GB:AC005027  251    16377  16377  A>G
              source        dbSNP  gnl|dbSNP|ss577365_allele
              source        dbSNP  gnl|dbSNP|ss588511_allele
              consequence  GB:AC005027_3526170CD1    252     5'
  GIF  INHBA-genomic-fwd.gif


IRS1
Full name : Insulin receptor substrate 1
Link : IRS1_link_cdna
  Subsequence      EM:S62539     1      5828   #254
  CDS  EM:S62539.1  3729 bp          #255
      ORF    1021   4749
  Allele      EM:S62539     254    3388   3388   A>G
              source        isSNP  SNP00067005
              consequence  EM:S62539.1  255    Missense     790-790      R>C
  Allele      EM:S62539     254    3887   3887   A>G
              source        isSNP  SNP00114530
              consequence  EM:S62539.1  255    Missense     956-956      E>G
  Allele      EM:S62539     254    5156   5156   G>T
              source        isSNP  SNP00067006
              consequence  EM:S62539.1  255    3'
  GIF  IRS1-cdna-fwd.gif
Link : IRS1_link_genomic
  Subsequence      EM:S85963     100    6251   #256
  Subsequence      IRS1_cds.1    680    4411   #257
  Subsequence      IRS1_mrna_build.1  100    4432   #258
  CDS  IRS1_cds.1     3732 bp      1 exon       #257
      exon   680    4411
  mRNA        IRS1_mrna_build.1  4333 bp      1 exon       #258
      exon   100    4432
```

TABLE 1 (Cont.)

Allele	EM:S85963	256	850	850	A>G		
	source	wetSNP		EM:S85963.v850.C>T			
	consequence	IRS1_cds.1		257	Silent	90-90	D
Allele	EM:S85963	256	1285	1285	A>G		
	source	wetSNP		EM:S85963.v1285.G>A			
	consequence	IRS1_cds.1		257	Silent	235-235	G
Allele	EM:S85963	256	1783	1783	A>G		
	source	wetSNP		EM:S85963.v1783.T>C			
	consequence	IRS1_cds.1		257	Silent	401-401	H
Allele	EM:S85963	256	2023	2023	A>G		
	source	wetSNP		EM:S85963.v2023.C>T			
	consequence	IRS1_cds.1		257	Silent	481-481	N
Allele	EM:S85963	256	2117	2117	C>G		
	source	wetSNP		EM:S85963.v2117.G>C			
	consequence	IRS1_cds.1		257	Missense	513-513	A>P
Allele	EM:S85963	256	2697	2697	A>G		
	source	wetSNP		EM:S85963.v2697.G>A			
	consequence	IRS1_cds.1		257	Missense	706-706	G>D
Allele	EM:S85963	256	2941	2941	A>G		
	source	wetSNP		EM:S85963.v2941.T>C			
	consequence	IRS1_cds.1		257	Silent	787-787	H
Allele	EM:S85963	256	2951	2951	A>G		
	source	isSNP SNP00067005					
	consequence	IRS1_cds.1		257	Missense	791-791	R>C
Allele	EM:S85963	256	2995	2995	A>G		
	source	wetSNP		EM:S85963.v2995.A>G			
	consequence	IRS1_cds.1		257	Silent	805-805	A
Allele	EM:S85963	256	3035	3035	C>G		
	source	wetSNP		EM:S85963.v3035.G>C			
	consequence	IRS1_cds.1		257	Missense	819-819	G>R
Allele	EM:S85963	256	3262	3262	C>G		
	source	wetSNP		EM:S85963.v3262.G>C			
	consequence	IRS1_cds.1		257	Silent	894-894	P
Allele	EM:S85963	256	3349	3349	A>G		
	source	wetSNP		EM:S85963.v3349.G>A			
	consequence	IRS1_cds.1		257	Silent	923-923	R
Allele	EM:S85963	256	3450	3450	A>G		
	source	isSNP SNP00114530					
	consequence	IRS1_cds.1		257	Missense	957-957	E>G
Allele	EM:S85963	256	3494	3494	A>G		
	source	wetSNP		EM:S85963.v3494.G>A			
	consequence	IRS1_cds.1		257	Missense	972-972	G>R
Allele	EM:S85963	256	4053	4053	A>G		
	source	wetSNP		EM:S85963.v4053.G>A			
	consequence	IRS1_cds.1		257	Missense	1158-1158	G>E

GIF IRS1-genomic-fwd.gif

JUN
Full name : v-jun avian sarcoma virus 17 oncogene homolog
Link : JUN_link_genomic

Subsequence	JUN_cds.1	9468	8473	#259	
Subsequence	GB:AL136985_1	1	151212	#260	
Subsequence	JUN_mrna_build.1	9468	8473	#261	

TABLE 1 (Cont.)

```
  CDS JUN_cds.1     996 bp       1 exon       #259
      exon   9468   8473
  mRNA          JUN_mrna_build.1   996 bp       1 exon       #261
      exon   9468   8473
  GIF JUN-genomic-rev.gif


KJ_OA11
Full name : KIAA1253
Link : FL_2135776_link_cdna
  Subsequence       FN:2135776CB1      1      3129   #262
  CDS FN:2135776CB1.1     1197 bp            #263
      ORF    256    1452
  Allele       FN:2135776CB1       262    59     59     C>G
               source        isSNP SNP00100733
               consequence  FN:2135776CB1.1     263    5'
  Allele       FN:2135776CB1       262    1352   1352   A>G
               source        isSNP SNP00116557
               consequence  FN:2135776CB1.1     263    Missense     366-366      Q>R
  Allele       FN:2135776CB1       262    1477   1477   A>G
               source        isSNP SNP00042286
               consequence  FN:2135776CB1.1     263    3'
  Allele       FN:2135776CB1       262    1489   1489   A>G
               source        isSNP SNP00042287
               consequence  FN:2135776CB1.1     263    3'
  Allele       FN:2135776CB1       262    1667   1667   A>G
               source        isSNP SNP00011480
               consequence  FN:2135776CB1.1     263    3'
  Allele       FN:2135776CB1       262    1710   1710   A>G
               source        isSNP SNP00011481
               consequence  FN:2135776CB1.1     263    3'
  Allele       FN:2135776CB1       262    1838   1838   A>G
               source        isSNP SNP00011482
               consequence  FN:2135776CB1.1     263    3'
  Allele       FN:2135776CB1       262    2589   2589   A>G
               source        isSNP SNP00003671
               consequence  FN:2135776CB1.1     263    3'
  GIF KJ_OA11-cdna-fwd.gif
Link : FL_2135776_link_genomic
  Subsequence       GB:HS425C14  1      160203       #264
  Subsequence       GB:HS425C14_2135776CD1    55766  42255  #265
  Subsequence       FL_2135776_mrna_build.1   69012  40562  #266
  Subsequence       KJ_OA11_cds.1       55766  51052  #267
  CDS GB:HS425C14_2135776CD1    1197 bp      9 exons      #265
      exon   55766  55731
      exon   53861  53692
      exon   51441  51362
      exon   51118  50981
      exon   49268  49099
      exon   48965  48875
      exon   44476  44332
      exon   44215  43985
      exon   42390  42255
  mRNA         FL_2135776_mrna_build.1  3119 bp      10 exons      #266
```

TABLE 1 (Cont.)

```
     exon    69012 68910
     exon    55892 55731
     exon    53861 53692
     exon    51441 51362
     exon    51118 50981
     exon    49268 49099
     exon    48965 48875
     exon    44476 44332
     exon    44215 43985
     exon    42390 40562
CDS  KJ_OA11_cds.1       273 bp      3 exons      #267
     exon    55766 55731
     exon    53861 53692
     exon    51118 51052
Allele       GB:HS425C14  264    41092 41092  A>G
             source       isSNP  SNP00003671
             consequence  GB:HS425C14_2135776CD1    265    3'
             consequence  KJ_OA11_cds.1        267    3'
Allele       GB:HS425C14  264    41843 41843  A>G
             source       isSNP  SNP00011482
             consequence  GB:HS425C14_2135776CD1    265    3'
             consequence  KJ_OA11_cds.1        267    3'
Allele       GB:HS425C14  264    41971 41971  A>G
             source       isSNP  SNP00011481
             consequence  GB:HS425C14_2135776CD1    265    3'
             consequence  KJ_OA11_cds.1        267    3'
Allele       GB:HS425C14  264    42014 42014  A>G
             source       isSNP  SNP00011480
             consequence  GB:HS425C14_2135776CD1    265    3'
             consequence  KJ_OA11_cds.1        267    3'
Allele       GB:HS425C14  264    42192 42192  A>G
             source       isSNP  SNP00042287
             consequence  GB:HS425C14_2135776CD1    265    3'
             consequence  KJ_OA11_cds.1        267    3'
Allele       GB:HS425C14  264    42204 42204  A>G
             source       isSNP  SNP00042286
             source       wetSNP       GB:HS425C14.v42204.G>A
             consequence  GB:HS425C14_2135776CD1    265    3'
             consequence  KJ_OA11_cds.1        267    3'
Allele       GB:HS425C14  264    42294 42294  C>G
             source       wetSNP       GB:HS425C14.v42294.G>C
             source       wetSNP       GB:HS425C14.v42294.G>C
             consequence  GB:HS425C14_2135776CD1    265    Silent       386-386
     L
             consequence  KJ_OA11_cds.1        267    3'
Allele       GB:HS425C14  264    42329 42329  A>G
             source       isSNP  SNP00116557
             consequence  GB:HS425C14_2135776CD1    265    Missense     375-375
     S>G
             consequence  KJ_OA11_cds.1        267    3'
Allele       GB:HS425C14  264    44297 44297  A>G
             source       wetSNP       GB:HS425C14.v44297.T>C
             consequence  GB:HS425C14_2135776CD1    265    Intron
             consequence  KJ_OA11_cds.1        267    3'
Allele       GB:HS425C14  264    55697 55697  A>G
```

TABLE 1 (Cont.)

```
          source        wetSNP      GB:HS425C14.v55697.C>T
          consequence   GB:HS425C14_2135776CD1    265    Intron
          consequence   KJ_OA11_cds.1       267    Intron
  Allele  GB:HS425C14   264    68954  68954  C>G
          source        isSNP  SNP00100733
          consequence   GB:HS425C14_2135776CD1    265    5'
          consequence   KJ_OA11_cds.1       267    5'
  GIF  KJ_OA11-genomic-rev.gif


KJ_OA2
Link : KJ_OA2_link_cdna
  Subsequence       LG:244552.16       1      1825   #268
  Allele     LG:244552.16        268    1476   1476   G>T
             source        isSNP  SNP00098862


KJ_OA21
Full name : FL project 2027624
Link : FL_2027624_link_cdna
  Subsequence       FN:2027624CB1       1      2173   #269
  CDS  FN:2027624CB1.1     1734 bp             #270
      ORF    4      1737
  Allele     FN:2027624CB1       269    881    881    C>G
             source        isSNP  SNP00106459
             consequence  FN:2027624CB1.1     270    Missense     293-293      T>R
  Allele     FN:2027624CB1       269    971    971    A>G
             source        isSNP  SNP00075286
             consequence  FN:2027624CB1.1     270    Missense     323-323      T>I
  Allele     FN:2027624CB1       269    1092   1092   C>G
             source        isSNP  SNP00106460
             consequence  FN:2027624CB1.1     270    Silent       363-363      L
  Allele     FN:2027624CB1       269    1254   1254   A>G
             source        isSNP  SNP00075287
             consequence  FN:2027624CB1.1     270    Silent       417-417      Q
  Allele     FN:2027624CB1       269    1374   1374   A>G
             source        isSNP  SNP00009699
             consequence  FN:2027624CB1.1     270    Silent       457-457      T
  Allele     FN:2027624CB1       269    1392   1392   A>G
             source        isSNP  SNP00097916
             consequence  FN:2027624CB1.1     270    Silent       463-463      G
  Allele     FN:2027624CB1       269    1623   1623   A>G
             source        isSNP  SNP00009700
             consequence  FN:2027624CB1.1     270    Silent       540-540      Y
  GIF  KJ_OA21-cdna-fwd.gif
Link : FL_1250708_link_genomic
  Subsequence       GB:HS453C12  1      147620       #271
  Subsequence       GB:HS453C12_1394592CD1    87967  109084       #272
  Subsequence       GB:HS453C12_2027624CD1    20194  10528  #273
  Subsequence       FL_1394592_mrna_build.1   87945  110578       #274
  Subsequence       FL_2027624_mrna_build.1   20197  6152   #275
  Subsequence       OA21_cds.1    20194  17050  #276
  mRNA       FL_2027624_mrna_build.1   2172 bp      13 exons      #275
```

TABLE 1 (Cont.)

```
     exon   20197  20008
     exon   19834  19657
     exon   17499  17372
     exon   17056  16956
     exon   16847  16761
     exon   16215  16128
     exon   16019  15922
     exon   15823  15658
     exon   14968  14768
     exon   12135  11970
     exon   11855  11772
     exon   10777  10110
     exon   6168   6152
CDS  OA21_cds.1    372 bp       3 exons      #276
     exon   20194  20008
     exon   19834  19657
     exon   17056  17050
CDS  GB:HS453C12_2027624CD1    1734 bp      12 exons     #273
     exon   20194  20008
     exon   19834  19657
     exon   17499  17372
     exon   17056  16956
     exon   16847  16761
     exon   16215  16128
     exon   16019  15922
     exon   15823  15658
     exon   14968  14768
     exon   12135  11970
     exon   11855  11772
     exon   10777  10528
Allele      GB:HS453C12  271    10642  10642  A>G
            source       isSNP  SNP00009700
            source       wetSNP        GB:HS453C12.v10642.A>G
            source       wetSNP        GB:HS453C12.v10642.A>G
            consequence  OA21_cds.1    276    3'
            consequence  GB:HS453C12_2027624CD1    273    Silent       540-540
     Y
Allele      GB:HS453C12  271    11206  11206  A>G
            source       dbSNP  gnl|dbSNP|ss979258_allele
            consequence  OA21_cds.1    276    3'
            consequence  GB:HS453C12_2027624CD1    273    Intron
Allele      GB:HS453C12  271    11999  11999  A>G
            source       isSNP  SNP00009699
            source       wetSNP        GB:HS453C12.v11999.C>T
            source       wetSNP        GB:HS453C12.v11999.C>T
            consequence  OA21_cds.1    276    3'
            consequence  GB:HS453C12_2027624CD1    273    Silent       457-457
     T
Allele      GB:HS453C12  271    13494  13494  A>G
            source       isSNP  SNP00095042
            consequence  OA21_cds.1    276    3'
            consequence  GB:HS453C12_2027624CD1    273    Intron
Allele      GB:HS453C12  271    14913  14913  C>G
            source       isSNP  SNP00106460
            consequence  OA21_cds.1    276    3'
```

TABLE 1 (Cont.)

```
          consequence  GB:HS453C12_2027624CD1    273    Silent       363-363
     L
  Allele      GB:HS453C12  271     15723  15723  A>G
              source       isSNP  SNP00075286
              consequence  OA21_cds.1    276     3'
              consequence  GB:HS453C12_2027624CD1    273    Missense     323-323
     T>I
  GIF  KJ_OA21-genomic-rev.gif


KJ_OA29
Link : KJ_OA29_link_cdna
  Subsequence       LG:199489.1  1       3318    #277
  Allele      LG:199489.1  277     544     544     A>G
              source       isSNP  SNP00005297
  Allele      LG:199489.1  277     695     695     A>G
              source       isSNP  SNP00121995
  Allele      LG:199489.1  277     971     971     A>G
              source       isSNP  SNP00047679
  Allele      LG:199489.1  277     1312    1312    A>G
              source       isSNP  SNP00005298
  Allele      LG:199489.1  277     1445    1445    A>G
              source       isSNP  SNP00027647
  Allele      LG:199489.1  277     2370    2370    A>G
              source       isSNP  SNP00005297
  Allele      LG:199489.1  277     2521    2521    A>G
              source       isSNP  SNP00121995
  Allele      LG:199489.1  277     2797    2797    A>G
              source       isSNP  SNP00047679
  Allele      LG:199489.1  277     3138    3138    A>G
              source       isSNP  SNP00005298
  Allele      LG:199489.1  277     3271    3271    A>G
              source       isSNP  SNP00027647


KJ_OA3
Link : KJ_OA3_link_cdna
  Subsequence       LG:153511.1  1       1628    #278
  Allele      LG:153511.1  278     395     395     A>G
              source       isSNP  SNP00003503
  Allele      LG:153511.1  278     1101    1101    A>G
              source       isSNP  SNP00113687


KJ_OA31
Link : KJ_OA31_link_cdna
  Subsequence       LG:200972.2  1       2192    #279
  Allele      LG:200972.2  279     366     366     C>G
              source       isSNP  SNP00099556
  Allele      LG:200972.2  279     836     836     A>G
              source       isSNP  SNP00015954
  Allele      LG:200972.2  279     1037    1037    A>G
```

TABLE 1 (Cont.)

```
            source       isSNP  SNP00015955
  Allele    LG:200972.2  279    1361   1361   A>G
            source       isSNP  SNP00000598
  Allele    LG:200972.2  279    1697   1697   A>G
            source       isSNP  SNP00000599
  Allele    LG:200972.2  279    1975   1975   A>G
            source       isSNP  SNP00067907
  Allele    LG:200972.2  279    2027   2027   A>G
            source       isSNP  SNP00067908
```

```
KJ_OA33
Full name : cardiotrophin-like cytokine
Link : FL_1676240_link_genomic
  Subsequence     GB:AC005849_1      1        169144      #280
  Subsequence     KJ_OA33_cds.1      151862   143455      #281
  Subsequence     KJ_OA33_mrna_build.1      151907      142489      #282
  CDS  KJ_OA33_cds.1      678 bp     3 exons     #281
       exon  151862       151847
       exon  145945       145779
       exon  143949       143455
  mRNA       KJ_OA33_mrna_build.1     1689 bp     3 exons     #282
       exon  151907       151847
       exon  145945       145779
       exon  143949       142489
  GIF  KJ_OA33-genomic-rev.gif
```

```
KJ_OA39
Link : KJ_OA39_link_cdna
  Subsequence     LG:293953.1  1     940    #283
  Allele     LG:293953.1  283    679    679    G>T
             source       isSNP  SNP00110603
```

```
KJ_OA6
Full name : FL project 2840746
Link : FL_818498_link_genomic
  Subsequence     GB:AC005598  1      190000      #284
  Subsequence     GB:AC005598_2840746CD1    132700      133368      #285
  Subsequence     FL_2840746_mrna_build.1   132672      135584      #286
  CDS  GB:AC005598_2840746CD1     669 bp     1 exon      #285
       exon  132700       133368
  mRNA       FL_2840746_mrna_build.1  1087 bp     2 exons     #286
       exon  132672       133391
       exon  135218       135584
  Allele     GB:AC005598  284    132689      132689      A>G
             source       isSNP  SNP00005520
             consequence  GB:AC005598_2840746CD1    285    5'
  Allele     GB:AC005598  284    132843      132843      A>G
             source       wetSNP       GB:AC005598.v132843.C>T
             consequence  GB:AC005598_2840746CD1    285    Silent      48-48  S
```

TABLE 1 (Cont.)

```
Allele       GB:AC005598 284   132878        132878       A>G
             source      wetSNP      GB:AC005598.v132878.G>A
             consequence GB:AC005598_2840746CD1   285   Missense    60-60  R>H
Allele       GB:AC005598 284   132951        132951       A>G
             source      wetSNP      GB:AC005598.v132951.C>T
             consequence GB:AC005598_2840746CD1   285   Silent      84-84  F
Allele       GB:AC005598 284   132967        132967       A>G
             source      wetSNP      GB:AC005598.v132967.C>T
             consequence GB:AC005598_2840746CD1   285   Missense    90-90  P>S
Allele       GB:AC005598 284   133103        133103       G>T
             source      wetSNP      GB:AC005598.v133103.G>T
             consequence GB:AC005598_2840746CD1   285   Missense    135-135
    G>V
Allele       GB:AC005598 284   133481        133481       A>G
             source      wetSNP      GB:AC005598.v133481.C>T
             consequence GB:AC005598_2840746CD1   285   3'
GIF KJ_OA6-genomic-fwd.gif


KJ_oagba3
Link : KJ_oagba3_link_cdna
  Subsequence      LG:215642.2 1      2849   #287
  Allele      LG:215642.2 287    1475   1475   A>G
              source      isSNP  SNP00041601
  Allele      LG:215642.2 287    1963   1963   A>G
              source      isSNP  SNP00010951


LIF
Full name : leukemia inhibitory factor
Link : LIF_link_cdna
  Subsequence      GB:LIF        1      3848   #288
  CDS  GB:LIF.1    609  bp              #289
       ORF    45      653
  Allele      GB:LIF        288    1183   1183   G>T
              source        isSNP  SNP00036337
              consequence   GB:LIF.1      289    3'
  Allele      GB:LIF        288    1572   1572   A>G
              source        isSNP  SNP00099092
              consequence   GB:LIF.1      289    3'
  Allele      GB:LIF        288    1996   1996   C>G
              source        isSNP  SNP00099093
              consequence   GB:LIF.1      289    3'
  Allele      GB:LIF        288    2062   2062   G>T
              source        isSNP  SNP00099094
              consequence   GB:LIF.1      289    3'
  Allele      GB:LIF        288    2404   2404   A>G
              source        isSNP  SNP00099095
              consequence   GB:LIF.1      289    3'
  Allele      GB:LIF        288    3156   3156   A>G
              source        isSNP  SNP00036338
              consequence   GB:LIF.1      289    3'
  Allele      GB:LIF        288    3582   3582   A>G
```

TABLE 1 (Cont.)

```
            source        isSNP SNP00008778
            consequence   GB:LIF.1     289    3'
     GIF LIF-cdna-fwd.gif
    Link : OSM_link_genomic
     Subsequence       GB:AC004264  1       47188  #290
     Subsequence       LIF_cds.1     11398  8354   #291
     Subsequence       LIF_mrna_build.1    11442  5156   #292
     CDS  LIF_cds.1    609 bp        3 exons       #291
          exon   11398  11380
          exon   9636   9458
          exon   8764   8354
     mRNA         LIF_mrna_build.1   3851 bp      3 exons       #292
          exon   11442  11380
          exon   9636   9458
          exon   8764   5156
     Allele       GB:AC004264  290    5420   5420   A>G
                  source       isSNP  SNP00008778
                  consequence  LIF_cds.1     291    3'
     Allele       GB:AC004264  290    5846   5846   A>G
                  source       isSNP  SNP00036338
                  consequence  LIF_cds.1     291    3'
     Allele       GB:AC004264  290    6598   6598   A>G
                  source       isSNP  SNP00099095
                  consequence  LIF_cds.1     291    3'
     Allele       GB:AC004264  290    6940   6940   G>T
                  source       isSNP  SNP00099094
                  consequence  LIF_cds.1     291    3'
     Allele       GB:AC004264  290    7006   7006   C>G
                  source       isSNP  SNP00099093
                  consequence  LIF_cds.1     291    3'
     Allele       GB:AC004264  290    7435   7435   A>G
                  source       isSNP  SNP00099092
                  consequence  LIF_cds.1     291    3'
     Allele       GB:AC004264  290    7824   7824   G>T
                  source       isSNP  SNP00036337
                  consequence  LIF_cds.1     291    3'
     GIF  LIF-genomic-rev.gif


LUM
Full name : lumican
Link : FL_2676170_link_genomic
     Subsequence       GB:AC007115_1        1      180821        #293
     Subsequence       GB:AC007115_1_3128106CD1          87417  92234  #294
     Subsequence       FL_3128106_mrna_build.1  84719  92839  #295
     mRNA        FL_3128106_mrna_build.1  1926 bp      3 exons       #295
          exon   84719  84998
          exon   87396  88278
          exon   92077  92839
     CDS  GB:AC007115_1_3128106CD1        1020 bp      2 exons       #294
          exon   87417  88278
          exon   92077  92234
     Allele      GB:AC007115_1       293    89050  89050  A>G
                 source        dbSNP  gnl|dbSNP|ss852530_allele
```

TABLE 1 (Cont.)

```
          source      dbSNP gnl|dbSNP|ss897123_allele
          consequence GB:AC007115_1_3128106CD1      294    Intron
  Allele  GB:AC007115_1      293    89249 89249 A>G
          source      dbSNP gnl|dbSNP|ss855039_allele
          consequence GB:AC007115_1_3128106CD1      294    Intron
  GIF LUM-genomic-fwd.gif


METTL1
Full name : methyltransferase-like 1
Link : METTL1_link_cdna
  Subsequence      GB:Y18643_1  1      1292   #296
  CDS GB:Y18643_1.1      831 bp           #297
      ORF   49    879
  Allele    GB:Y18643_1  296   345   345   A>G
            source       isSNP SNP00098761
            consequence  GB:Y18643_1.1     297   Silent      99-99 P
  Allele    GB:Y18643_1  296   919   919   A>G
            source       isSNP SNP00003825
            consequence  GB:Y18643_1.1     297   3'
  GIF METTL1-cdna-fwd.gif


MMP1
Full name : matrix metalloproteinase 1
Link : MMP1_link_cdna
  Subsequence      EM:HSCOLL1   1      1970   #298
  Allele    EM:HSCOLL1   298   383   383   A>G
            source       isSNP SNP00009627
  Allele    EM:HSCOLL1   298   714   714   A>G
            source       isSNP SNP00037857
  Allele    EM:HSCOLL1   298   745   745   A>G
            source       isSNP SNP00037858
  Allele    EM:HSCOLL1   298   1522  1522  A>G
            source       isSNP SNP00009628
  Allele    EM:HSCOLL1   298   1541  1541  A>G
            source       isSNP SNP00009629
  Allele    EM:HSCOLL1   298   1662  1662  A>G
            source       isSNP SNP00009630
  Allele    EM:HSCOLL1   298   1747  1747  A>G
            source       isSNP SNP00009631
Link : MMP1_link_genomic
  Subsequence      GB:HSU78045  1      81826  #299
  Subsequence      MMP1_cds.1   11905  4225   #300
  Subsequence      MMP1_mrna_build.1  11973  3733   #301
  CDS MMP1_cds.1   1410 bp      10 exons      #300
      exon  11905 11801
      exon  11314 11070
      exon  10976 10828
      exon  10603 10478
      exon  9421  9266
      exon  9105  8988
      exon  6551  6418
```

TABLE 1 (Cont.)

exon	5308	5146					
exon	4619	4516					
exon	4334	4225					
mRNA	MMP1_mrna_build.1	1970 bp		10 exons	#301		
exon	11973	11801					
exon	11314	11070					
exon	10976	10828					
exon	10603	10478					
exon	9421	9266					
exon	9105	8988					
exon	6551	6418					
exon	5308	5146					
exon	4619	4516					
exon	4334	3733					
Allele	GB:HSU78045	299	3956	3956	A>G		
	source	isSNP	SNP00009631				
	consequence	MMP1_cds.1		300	3'		
Allele	GB:HSU78045	299	4041	4041	A>G		
	source	isSNP	SNP00009630				
	consequence	MMP1_cds.1		300	3'		
Allele	GB:HSU78045	299	4162	4162	A>G		
	source	isSNP	SNP00009629				
	consequence	MMP1_cds.1		300	3'		
Allele	GB:HSU78045	299	4181	4181	A>G		
	source	isSNP	SNP00009628				
	consequence	MMP1_cds.1		300	3'		
Allele	GB:HSU78045	299	4517	4517	A>G		
	source	wetSNP		GB:HSU78045.v4517.A>G			
	consequence	MMP1_cds.1		300	Silent	433-433	D
Allele	GB:HSU78045	299	4661	4664	CATG>CG		
	source	wetSNP		GB:HSU78045.v4661.CATG>CG			
	consequence	MMP1_cds.1		300	Intron		
Allele	GB:HSU78045	299	4677	4677	A>G		
	source	wetSNP		GB:HSU78045.v4677.G>A			
	consequence	MMP1_cds.1		300	Intron		
Allele	GB:HSU78045	299	5198	5198	A>G		
	source	wetSNP		GB:HSU78045.v5198.A>G			
	consequence	MMP1_cds.1		300	Missense	382-382	S>P
Allele	GB:HSU78045	299	6586	6586	A>G		
	source	wetSNP		GB:HSU78045.v6586.T>C			
	consequence	MMP1_cds.1		300	Intron		
Allele	GB:HSU78045	299	9056	9056	A>G		
	source	wetSNP		GB:HSU78045.v9056.C>T			
	consequence	MMP1_cds.1		300	Silent	277-277	A
Allele	GB:HSU78045	299	9120	9120	A>G		
	source	wetSNP		GB:HSU78045.v9120.A>G			
	consequence	MMP1_cds.1		300	Intron		
Allele	GB:HSU78045	299	9126	9126	A>G		
	source	wetSNP		GB:HSU78045.v9126.G>A			
	consequence	MMP1_cds.1		300	Intron		
Allele	GB:HSU78045	299	9205	9205	A>G		
	source	wetSNP		GB:HSU78045.v9205.T>C			
	consequence	MMP1_cds.1		300	Intron		
Allele	GB:HSU78045	299	9247	9247	A>G		
	source	wetSNP		GB:HSU78045.v9247.T>C			

TABLE 1 (Cont.)

```
            consequence  MMP1_cds.1   300    Intron
  Allele    GB:HSU78045  299    9365  9365   G>T
            source       wetSNP       GB:HSU78045.v9365.G>T
            consequence  MMP1_cds.1   300    Missense     228-228     H>N
  Allele    GB:HSU78045  299    9370  9370   A>G
            source       isSNP  SNP00037858
            consequence  MMP1_cds.1   300    Missense     226-226     L>P
  Allele    GB:HSU78045  299    11105 11105  A>G
            source       isSNP  SNP00009627
            source       wetSNP       GB:HSU78045.v11105.C>T
            consequence  MMP1_cds.1   300    Silent       105-105     G
  GIF  MMP1-genomic-rev.gif


MMP13
Full name : MMP13
Link : MMP13_link_genomic
  Subsequence       MMP13_cds.1  141623       159614      #302
  Subsequence       GB:AP000789_1      1      201766      #303
  CDS  MMP13_cds.1  957 bp       7 exons      #302
      exon   141629       141779
      exon   141956       142081
      exon   144063       144224
      exon   146009       146126
      exon   147078       147211
      exon   157208       157367
      exon   159509       159614
  Allele    GB:AP000789_1      303    141614       141614      C>G
            source       wetSNP       GB:AP000789_1.v141614.C>G
            consequence  MMP13_cds.1  302    5'
  Allele    GB:AP000789_1      303    141875       141875      G>T
            source       wetSNP       GB:AP000789_1.v141875.C>A
            consequence  MMP13_cds.1  302    Intron
  Allele    GB:AP000789_1      303    147095       147095      A>G
            source       wetSNP       GB:AP000789_1.v147095.A>G
            consequence  MMP13_cds.1  302    Missense     192-192     H>R
  Allele    GB:AP000789_1      303    157231       157231      C>G
            source       wetSNP       GB:AP000789_1.v157231.G>C
            consequence  MMP13_cds.1  302    Missense     239-239     G>R
  Allele    GB:AP000789_1      303    157325       157325      A>G
            source       wetSNP       GB:AP000789_1.v157325.A>G
            consequence  MMP13_cds.1  302    Missense     270-270     D>G
  Allele    GB:AP000789_1      303    159631       159631      A>G
            source       wetSNP       GB:AP000789_1.v159631.C>T
            consequence  MMP13_cds.1  302    3'
  Allele    GB:AP000789_1      303    159644       159644      C>G
            source       wetSNP       GB:AP000789_1.v159644.G>C
            consequence  MMP13_cds.1  302    3'
  GIF  MMP13-genomic-fwd.gif


MMP14
Full name : MMP14
```

TABLE 1 (Cont.)

```
Link : MMP14_link_cdna
  Subsequence      GB:HUMMTMMP  1      3403   #304
  CDS  GB:HUMMTMMP.1       1749 bp             #305
      ORF    112    1860
  Allele     GB:HUMMTMMP  304    133    133    A>G
             source       isSNP  SNP00107954
             consequence  GB:HUMMTMMP.1        305    Missense     8-8    S>P
  Allele     GB:HUMMTMMP  304    580    580    A>G
             source       isSNP  SNP00107955
             consequence  GB:HUMMTMMP.1        305    Silent       157-157      L
  Allele     GB:HUMMTMMP  304    888    888    C>G
             source       isSNP  SNP00093383
             consequence  GB:HUMMTMMP.1        305    Silent       259-259      P
  Allele     GB:HUMMTMMP  304    966    966    A>G
             source       isSNP  SNP00055171
             consequence  GB:HUMMTMMP.1        305    Silent       285-285      G
  Allele     GB:HUMMTMMP  304    1243   1243   A>G
             source       isSNP  SNP00107956
             consequence  GB:HUMMTMMP.1        305    Missense     378-378      K>E
  Allele     GB:HUMMTMMP  304    1264   1264   C>G
             source       isSNP  SNP00107957
             consequence  GB:HUMMTMMP.1        305    Missense     385-385      D>H
  Allele     GB:HUMMTMMP  304    1944   1944   A>G
             source       isSNP  SNP00060446
             consequence  GB:HUMMTMMP.1        305    3'
  GIF MMP14-cdna-fwd.gif
Link : MMP14_link_genomic
  Subsequence      MMP14_cds.1  132034       141254       #306
  Subsequence      GB:AL133448_3      1      173805       #307
  Subsequence      MMP14_mrna_build.1        131922       142801       #308
  CDS  MMP14_cds.1  1749 bp      10 exons      #306
      exon   132034       132141
      exon   136706       136854
      exon   137128       137250
      exon   137625       137932
      exon   138472       138633
      exon   138925       139085
      exon   139586       139724
      exon   139845       139995
      exon   140466       140581
      exon   140923       141254
  mRNA        MMP14_mrna_build.1       3408 bp      10 exons      #308
      exon   131922       132141
      exon   136706       136854
      exon   137128       137250
      exon   137625       137932
      exon   138472       138633
      exon   138925       139085
      exon   139586       139724
      exon   139845       139995
      exon   140466       140581
      exon   140923       142801
  Allele      GB:AL133448_3       307    132055        132055        A>G
             source       isSNP  SNP00107954
             consequence  MMP14_cds.1  306    Missense     8-8    P>S
```

TABLE 1 (Cont.)

```
Allele      GB:AL133448_3       307    137049        137051        TTA>TA
            source        wetSNP       GB:AL133448_3.v137049.TTA>TA
            consequence  MMP14_cds.1  306    Intron
Allele      GB:AL133448_3       307    137713        137713        A>G
            source        isSNP  SNP00107955
            consequence  MMP14_cds.1  306    Silent        157-157       L
Allele      GB:AL133448_3       307    138406        138406        A>G
            source        wetSNP       GB:AL133448_3.v138406.G>A
            consequence  MMP14_cds.1  306    Intron
Allele      GB:AL133448_3       307    138560        138560        C>G
            source        isSNP  SNP00093383
            source        wetSNP       GB:AL133448_3.v138560.C>G
            consequence  MMP14_cds.1  306    Silent        259-259       P
Allele      GB:AL133448_3       307    138653        138653        A>G
            source        wetSNP       GB:AL133448_3.v138653.G>A
            consequence  MMP14_cds.1  306    Intron
Allele      GB:AL133448_3       307    139639        139639        A>G
            source        wetSNP       GB:AL133448_3.v139639.G>A
            consequence  MMP14_cds.1  306    Missense      355-355       M>I
Allele      GB:AL133448_3       307    139981        139981        A>G
            source        wetSNP       GB:AL133448_3.v139981.C>T
            consequence  MMP14_cds.1  306    Silent        429-429       F
Allele      GB:AL133448_3       307    139986        139986        A>G
            source        wetSNP       GB:AL133448_3.v139986.G>A
            consequence  MMP14_cds.1  306    Missense      431-431       R>H
Allele      GB:AL133448_3       307    141337        141337        A>G
            source        isSNP  SNP00060446
            consequence  MMP14_cds.1  306    3'
 GIF  MMP14-genomic-fwd.gif


MMP2
Link  :  MMP2_link_cdna
  Subsequence        GB:HSMMPM2    1       3530     #309
  CDS  GB:HSMMPM2.1           2010  bp              #310
       ORF    49      2058
  Allele      GB:HSMMPM2    309    681    681     A>G
              source        isSNP  SNP00100004
              consequence   GB:HSMMPM2.1          310    Silent        211-211       P
  Allele      GB:HSMMPM2    309    1835   1835    A>G
              source        isSNP  SNP00100005
              consequence   GB:HSMMPM2.1          310    Missense      596-596       D>G
  Allele      GB:HSMMPM2    309    1851   1851    G>T
              source        isSNP  SNP00075435
              consequence   GB:HSMMPM2.1          310    Missense      601-601       F>L
  Allele      GB:HSMMPM2    309    2717   2717    A>G
              source        isSNP  SNP00024650
              consequence   GB:HSMMPM2.1          310    3'
  Allele      GB:HSMMPM2    309    2922   2922    C>G
              source        isSNP  SNP00024651
              consequence   GB:HSMMPM2.1          310    3'
  GIF  MMP2-cdna-fwd.gif
Link  :  MMP2_link_genomic
  Subsequence        MMP2_cds.1    175558        156463        #311
```

TABLE 1 (Cont.)

```
Subsequence     GB:AC012182_3       1       190117       #312
Subsequence     MMP2_mrna_build.1  175606       155007       #313
CDS  MMP2_cds.1   2010 bp      10 exons     #311
     exon  175558        175397
     exon  164437        164289
     exon  163643        163515
     exon  162034        161727
     exon  161372        161211
     exon  160292        160039
     exon  159678        159540
     exon  158699        158549
     exon  158397        158282
     exon  156902        156463
mRNA       MMP2_mrna_build.1  3514 bp      10 exons     #313
     exon  175606        175397
     exon  164437        164289
     exon  163643        163515
     exon  162034        161727
     exon  161372        161211
     exon  160292        160039
     exon  159678        159540
     exon  158699        158549
     exon  158397        158282
     exon  156902        155007
Allele     GB:AC012182_3       312    155598       155598       C>G
           source       isSNP  SNP00024651
           consequence  MMP2_cds.1   311    3'
Allele     GB:AC012182_3       312    155804       155804       A>G
           source       isSNP  SNP00024650
           consequence  MMP2_cds.1   311    3'
Allele     GB:AC012182_3       312    156670       156670       G>T
           source       isSNP  SNP00075435
           consequence  MMP2_cds.1   311    Missense     601-601      F>L
Allele     GB:AC012182_3       312    156686       156686       A>G
           source       isSNP  SNP00100005
           consequence  MMP2_cds.1   311    Missense     596-596      D>G
Allele     GB:AC012182_3       312    161842       161842       A>G
           source       isSNP  SNP00100004
           consequence  MMP2_cds.1   311    Silent       211-211      P
Allele     GB:AC012182_3       312    163660       163660       A>G
           source       wetSNP       GB:AC012182_3.v163660.G>A
           consequence  MMP2_cds.1   311    Intron
GIF MMP2-genomic-rev.gif


MMP3
Full name : matrix metalloproteinase 3
Link : MMP3_link_cdna
 Subsequence      EM:HSSTROMR  1      1801    #314
 Allele     EM:HSSTROMR  314    331    331    A>G
            source       isSNP  SNP00011525
 Allele     EM:HSSTROMR  314    382    382    A>G
            source       isSNP  SNP00113489
 Allele     EM:HSSTROMR  314    713    713    A>G
```

TABLE 1 (Cont.)

```
            source      isSNP  SNP00015044
  Allele    EM:HSSTROMR 314    976    976    A>G
            source      isSNP  SNP00054705
  Allele    EM:HSSTROMR 314    1129   1129   A>G
            source      isSNP  SNP00011527
Link : MMP3_link_genomic
  Subsequence     EM:HSU78045  100    81925 #315
  Subsequence     MMP3_link_cds.1     57437 50020 #316
  Subsequence     MMP3_mrna_build.1   57480 49696 #317
  CDS MMP3_link_cds.1    1434 bp     10 exons    #316
      exon  57437 57333
      exon  56806 56562
      exon  56469 56321
      exon  56182 56057
      exon  54487 54323
      exon  54146 54002
      exon  53137 53004
      exon  52604 52445
      exon  51295 51192
      exon  50120 50020
  mRNA      MMP3_mrna_build.1  1801 bp     10 exons    #317
      exon  57480 57333
      exon  56806 56562
      exon  56469 56321
      exon  56182 56057
      exon  54487 54323
      exon  54146 54002
      exon  53137 53004
      exon  52604 52445
      exon  51295 51192
      exon  50120 49696
  Allele    EM:HSU78045 315    52375 52375 A>G
            source      wetSNP      EM:HSU78045.v52375.T>C
            consequence MMP3_link_cds.1    316    Silent     400-400     T
  Allele    EM:HSU78045 315    52411 52411 A>G
            source      wetSNP      EM:HSU78045.v52411.G>A
            consequence MMP3_link_cds.1    316    Silent     388-388     I
  Allele    EM:HSU78045 315    52489 52489 A>G
            source      wetSNP      EM:HSU78045.v52489.G>A
            consequence MMP3_link_cds.1    316    Silent     362-362     A
  Allele    EM:HSU78045 315    52527 52530 GAGT>GT
            source      wetSNP      EM:HSU78045.v52527.GAGT>GT
            consequence MMP3_link_cds.1    316    Intron
  Allele    EM:HSU78045 315    52586 52586 A>T
            source      wetSNP      EM:HSU78045.v52586.T>A
            consequence MMP3_link_cds.1    316    Intron
  Allele    EM:HSU78045 315    53771 53771 A>T
            source      wetSNP      EM:HSU78045.v53771.T>A
            consequence MMP3_link_cds.1    316    Intron
  Allele    EM:HSU78045 315    54077 54077 C>G
            source      wetSNP      EM:HSU78045.v54077.C>G
            consequence MMP3_link_cds.1    316    Intron
  Allele    EM:HSU78045 315    54187 54187 A>G
            source      wetSNP      EM:HSU78045.v54187.C>T
            consequence MMP3_link_cds.1    316    Intron
```

TABLE 1 (Cont.)

Allele EM:HSU78045 315 54402 54402 A>G
source wetSNP EM:HSU78045.v54402.C>T
consequence MMP3_link_cds.1 316 Intron
Allele EM:HSU78045 315 56119 56119 A>G
source wetSNP EM:HSU78045.v56119.C>T
consequence MMP3_link_cds.1 316 Intron
Allele EM:HSU78045 315 56507 56507 C>G
source wetSNP EM:HSU78045.v56507.G>C
consequence MMP3_link_cds.1 316 Silent 102-102 T
Allele EM:HSU78045 315 56525 56525 A>G
source isSNP SNP00011525
source wetSNP EM:HSU78045.v56525.G>A
consequence MMP3_link_cds.1 316 Silent 96-96 D
Allele EM:HSU78045 315 56680 56680 A>G
source wetSNP EM:HSU78045.v56680.C>T
consequence MMP3_link_cds.1 316 Missense 45-45 E>K
GIF MMP3-genomic-rev.gif

MMP9
Full name : matrix metalloproteinase 9
Link : MMP9_link_cdna
Subsequence FN:522678CB1 1 2348 #318
CDS FN:522678CB1.1 2124 bp #319
ORF 33 2156
Allele FN:522678CB1 318 308 308 A>G
source isSNP SNP00101082
consequence FN:522678CB1.1 319 Silent 92-92 K
Allele FN:522678CB1 318 413 413 A>G
source isSNP SNP00101083
consequence FN:522678CB1.1 319 Silent 127-127 N
Allele FN:522678CB1 318 534 534 A>G
source isSNP SNP00101084
consequence FN:522678CB1.1 319 Missense 168-168 I>V
Allele FN:522678CB1 318 591 591 A>G
source isSNP SNP00101085
consequence FN:522678CB1.1 319 Missense 187-187 L>F
Allele FN:522678CB1 318 719 719 A>G
source isSNP SNP00101086
consequence FN:522678CB1.1 319 Silent 229-229 A
Allele FN:522678CB1 318 748 748 A>G
source isSNP SNP00021346
consequence FN:522678CB1.1 319 Missense 239-239 R>H
Allele FN:522678CB1 318 868 868 A>G
source isSNP SNP00002987
consequence FN:522678CB1.1 319 Missense 279-279 Q>R
Allele FN:522678CB1 318 1604 1604 A>G
source isSNP SNP00021347
consequence FN:522678CB1.1 319 Silent 524-524 I
Allele FN:522678CB1 318 1853 1853 G>T
source isSNP SNP00002988
consequence FN:522678CB1.1 319 Silent 607-607 G
Allele FN:522678CB1 318 2159 2159 A>G
source isSNP SNP00062663

TABLE 1 (Cont.)

```
            consequence  FN:522678CB1.1       319    3'
   Allele   FN:522678CB1         318    2302  2302   A>G
            source       isSNP  SNP00021348
            consequence  FN:522678CB1.1       319    3'
   GIF MMP9-cdna-fwd.gif
 Link : MMP9_link_genomic
   Subsequence      GB:HUMIVCOL01       1      764    #320
   Subsequence      GB:HUMIVCOL02       865    1117   #321
   Subsequence      GB:HUMIVCOL03       1218   1386   #322
   Subsequence      GB:HUMIVCOL04       1487   1635   #323
   Subsequence      GB:HUMIVCOL05       1736   1929   #324
   Subsequence      GB:HUMIVCOL06       2030   2223   #325
   Subsequence      GB:HUMIVCOL07       2324   2520   #326
   Subsequence      GB:HUMIVCOL08       2621   2796   #327
   Subsequence      GB:HUMIVCOL09       2897   3196   #328
   Subsequence      GB:HUMIVCOL10       3297   3456   #329
   Subsequence      GB:HUMIVCOL11       3557   3727   #330
   Subsequence      GB:HUMIVCOL12       3828   3951   #331
   Subsequence      GB:HUMIVCOL13       4052   4371   #332
   Subsequence      MMP9_cds.1    619   4180   #333
   Subsequence      MMP9_mrna_build.1   587    4371   #334
   CDS  MMP9_cds.1    2124 bp     13 exons     #333
        exon   619    756
        exon   875    1107
        exon   1228   1376
        exon   1497   1625
        exon   1746   1919
        exon   2040   2213
        exon   2334   2510
        exon   2631   2786
        exon   2907   3186
        exon   3307   3446
        exon   3567   3717
        exon   3838   3941
        exon   4062   4180
   mRNA         MMP9_mrna_build.1  2348 bp     13 exons      #334
        exon   587    756
        exon   875    1107
        exon   1228   1376
        exon   1497   1625
        exon   1746   1919
        exon   2040   2213
        exon   2334   2510
        exon   2631   2786
        exon   2907   3186
        exon   3307   3446
        exon   3567   3717
        exon   3838   3941
        exon   4061   4371
   Allele       GB:HUMIVCOL01      320    677    677    A>G
                source        wetSNP       GB:HUMIVCOL01.v677.C>T
                consequence  MMP9_cds.1    333    Missense     20-20  A>V
   Allele       GB:HUMIVCOL02      321    148    148    A>G
                source        isSNP  SNP00101082
                consequence  MMP9_cds.1    333    Silent       92-92  K
```

TABLE 1 (Cont.)

```
  Allele      GB:HUMIVCOL04       323    49     49     A>G
              source       isSNP  SNP00101085
              consequence  MMP9_cds.1    333    Missense     187-187      L>F
  Allele      GB:HUMIVCOL05       324    48     48     A>G
              source       isSNP  SNP00101086
              consequence  MMP9_cds.1    333    Silent       229-229      A
  Allele      GB:HUMIVCOL05       324    77     77     A>G
              source       isSNP  SNP00021346
              consequence  MMP9_cds.1    333    Missense     239-239      R>H
  Allele      GB:HUMIVCOL09       328    252    252    A>G
              source       isSNP  SNP00021347
              consequence  MMP9_cds.1    333    Silent       524-524      I
  Allele      GB:HUMIVCOL11       330    81     81     G>T
              source       isSNP  SNP00002988
              consequence  MMP9_cds.1    333    Silent       607-607      G
  Allele      GB:HUMIVCOL13       332    87     87     A>G
              source       wetSNP        GB:HUMIVCOL13.v87.G>A
              consequence  MMP9_cds.1    333    Silent       694-694      V
  Allele      GB:HUMIVCOL13       332    132    132    A>G
              source       wetSNP        GB:HUMIVCOL13.v132.C>T
              consequence  MMP9_cds.1    333    3'
  Allele      GB:HUMIVCOL13       332    274    274    A>G
              source       isSNP  SNP00021348
              consequence  MMP9_cds.1    333    3'
  GIF MMP9-genomic-fwd.gif


MSF
Full name : megakaryocyte stimulating factor
Link : MSF_link_cdna
  Subsequence       GB:NM_005807        1      5041   #335
  CDS GB:NM_005807.1      4215 bp              #336
      ORF    34     4248
  Allele      GB:NM_005807        335    1011   1011   A>G
              source       isSNP  SNP00064566
              consequence  GB:NM_005807.1      336    Silent       326-326      K
  Allele      GB:NM_005807        335    2650   2650   A>G
              source       isSNP  SNP00108532
              consequence  GB:NM_005807.1      336    Missense     873-873      P>S
  Allele      GB:NM_005807        335    3171   3171   A>G
              source       isSNP  SNP00009620
              consequence  GB:NM_005807.1      336    Silent       1046-1046    P
  Allele      GB:NM_005807        335    4187   4187   A>G
              source       isSNP  SNP00061665
              consequence  GB:NM_005807.1      336    Missense     1385-1385    A>V
  Allele      GB:NM_005807        335    4760   4760   A>G
              source       isSNP  SNP00009621
              consequence  GB:NM_005807.1      336    3'
  GIF MSF-cdna-fwd.gif
Link : MSF_link_genomic
  Subsequence       MSF_cds.1     181003        197905        #337
  Subsequence       MSF_cds.2     181003        197905        #338
  Subsequence       MSF_cds.3     181003        197905        #339
  Subsequence       MSF_cds.4     181003        197905        #340
```

TABLE 1 (Cont.)

```
Subsequence     GB:AL133553_7       1        214019       #341
Subsequence     MSF_mrna_build.1    180982    198681       #342
CDS  MSF_cds.3    3936 bp     10 exons     #339
     exon   181003        181078
     exon   184218        184340
     exon   185719        185838
     exon   190445        193267
     exon   193920        193997
     exon   195161        195297
     exon   195567        195723
     exon   196302        196499
     exon   196896        197021
     exon   197808        197905
mRNA        MSF_mrna_build.1    5012 bp     12 exons     #342
     exon   180982        181078
     exon   184218        184340
     exon   185719        185838
     exon   188235        188384
     exon   188921        189049
     exon   190445        193267
     exon   193920        193997
     exon   195161        195297
     exon   195567        195723
     exon   196302        196499
     exon   196896        197021
     exon   197808        198681
CDS  MSF_cds.4    3813 bp     9 exons     #340
     exon   181003        181078
     exon   185719        185838
     exon   190445        193267
     exon   193920        193997
     exon   195161        195297
     exon   195567        195723
     exon   196302        196499
     exon   196896        197021
     exon   197808        197905
CDS  MSF_cds.1    4215 bp     12 exons     #337
     exon   181003        181078
     exon   184218        184340
     exon   185719        185838
     exon   188235        188384
     exon   188921        189049
     exon   190445        193267
     exon   193920        193997
     exon   195161        195297
     exon   195567        195723
     exon   196302        196499
     exon   196896        197021
     exon   197808        197905
CDS  MSF_cds.2    4092 bp     11 exons     #338
     exon   181003        181078
     exon   185719        185838
     exon   188235        188384
     exon   188921        189049
     exon   190445        193267
```

TABLE 1 (Cont.)

exon	193920	193997				
exon	195161	195297				
exon	195567	195723				
exon	196302	196499				
exon	196896	197021				
exon	197808	197905				
Allele	GB:AL133553_7	341	190505	190505	G>T	
	source	wetSNP	GB:AL133553_7.v190505.A>C			
	consequence	MSF_cds.3	339	Missense	127-127	D>A
	consequence	MSF_cds.4	340	Missense	86-86	D>A
	consequence	MSF_cds.1	337	Missense	220-220	D>A
	consequence	MSF_cds.2	338	Missense	179-179	D>A
Allele	GB:AL133553_7	341	190559	190559	A>G	
	source	wetSNP	GB:AL133553_7.v190559.C>T			
	consequence	MSF_cds.3	339	Missense	145-145	T>M
	consequence	MSF_cds.4	340	Missense	104-104	T>M
	consequence	MSF_cds.1	337	Missense	238-238	T>M
	consequence	MSF_cds.2	338	Missense	197-197	T>M
Allele	GB:AL133553_7	341	190755	190755	A>G	
	source	wetSNP	GB:AL133553_7.v190755.G>A			
	consequence	MSF_cds.3	339	Silent	210-210	K
	consequence	MSF_cds.4	340	Silent	169-169	K
	consequence	MSF_cds.1	337	Silent	303-303	K
	consequence	MSF_cds.2	338	Silent	262-262	K
Allele	GB:AL133553_7	341	190824	190824	A>G	
	source	isSNP	SNP00064566			
	consequence	MSF_cds.3	339	Silent	233-233	K
	consequence	MSF_cds.4	340	Silent	192-192	K
	consequence	MSF_cds.1	337	Silent	326-326	K
	consequence	MSF_cds.2	338	Silent	285-285	K
Allele	GB:AL133553_7	341	192463	192463	A>G	
	source	isSNP	SNP00108532			
	consequence	MSF_cds.3	339	Missense	780-780	P>S
	consequence	MSF_cds.4	340	Missense	739-739	P>S
	consequence	MSF_cds.1	337	Missense	873-873	P>S
	consequence	MSF_cds.2	338	Missense	832-832	P>S
Allele	GB:AL133553_7	341	192984	192984	A>G	
	source	isSNP	SNP00009620			
	consequence	MSF_cds.3	339	Silent	953-953	P
	consequence	MSF_cds.4	340	Silent	912-912	P
	consequence	MSF_cds.1	337	Silent	1046-1046	P
	consequence	MSF_cds.2	338	Silent	1005-1005	P
Allele	GB:AL133553_7	341	193235	193235	A>G	
	source	wetSNP	GB:AL133553_7.v193235.A>G			
	consequence	MSF_cds.3	339	Missense	1037-1037	N>S
	consequence	MSF_cds.4	340	Missense	996-996	N>S
	consequence	MSF_cds.1	337	Missense	1130-1130	N>S
	consequence	MSF_cds.2	338	Missense	1089-1089	N>S
Allele	GB:AL133553_7	341	193258	193258	A>G	
	source	wetSNP	GB:AL133553_7.v193258.A>G			
	consequence	MSF_cds.3	339	Missense	1045-1045	M>V
	consequence	MSF_cds.4	340	Missense	1004-1004	M>V
	consequence	MSF_cds.1	337	Missense	1138-1138	M>V
	consequence	MSF_cds.2	338	Missense	1097-1097	M>V
Allele	GB:AL133553_7	341	196691	196691	G>T	

TABLE 1 (Cont.)

```
              source        isSNP  SNP00023429
              consequence   MSF_cds.3     339     Intron
              consequence   MSF_cds.4     340     Intron
              consequence   MSF_cds.1     337     Intron
              consequence   MSF_cds.2     338     Intron
   Allele     GB:AL133553_7        341    197844         197844        A>G
              source        isSNP  SNP00061665
              consequence   MSF_cds.3     339     Missense      1292-1292     A>V
              consequence   MSF_cds.4     340     Missense      1251-1251     A>V
              consequence   MSF_cds.1     337     Missense      1385-1385     A>V
              consequence   MSF_cds.2     338     Missense      1344-1344     A>V
   Allele     GB:AL133553_7        341    198417         198417        A>G
              source        isSNP  SNP00009621
              consequence   MSF_cds.3     339     3'
              consequence   MSF_cds.4     340     3'
              consequence   MSF_cds.1     337     3'
              consequence   MSF_cds.2     338     3'
   GIF  MSF-genomic-fwd.gif


NCOR2
Full name : nuclear receptor co-repressor 2
Link : NCOR2_link_cdna
   Subsequence        GB:AF125672  1       8686    #343
   CDS  GB:AF125672.1          7524 bp             #344
        ORF     157     7680
   Allele       GB:AF125672  343    165     165    G>T
                source        isSNP  SNP00035702
                consequence   GB:AF125672.1         344     Silent        3-3    G
   Allele       GB:AF125672  343    618     618    A>G
                source        isSNP  SNP00105557
                consequence   GB:AF125672.1         344     Silent        154-154       P
   Allele       GB:AF125672  343    2859    2859   A>G
                source        isSNP  SNP00101011
                consequence   GB:AF125672.1         344     Silent        901-901       A
   Allele       GB:AF125672  343    4728    4728   A>G
                source        isSNP  SNP00075034
                consequence   GB:AF125672.1         344     Silent        1524-1524     G
   Allele       GB:AF125672  343    4749    4749   A>G
                source        isSNP  SNP00069757
                consequence   GB:AF125672.1         344     Silent        1531-1531     L
   Allele       GB:AF125672  343    4957    4957   A>G
                source        isSNP  SNP00101012
                consequence   GB:AF125672.1         344     Missense      1601-1601     Y>H
   Allele       GB:AF125672  343    5085    5085   A>G
                source        isSNP  SNP00075035
                consequence   GB:AF125672.1         344     Silent        1643-1643     R
   Allele       GB:AF125672  343    5100    5100   A>G
                source        isSNP  SNP00075036
                consequence   GB:AF125672.1         344     Silent        1648-1648     N
   Allele       GB:AF125672  343    5221    5221   A>G
                source        isSNP  SNP00012485
                consequence   GB:AF125672.1         344     Missense      1689-1689     T>A
   Allele       GB:AF125672  343    7405    7405   A>G
```

TABLE 1 (Cont.)

```
             source         isSNP  SNP00015859
             consequence    GB:AF125672.1       344    Missense      2417-2417     P>S
   Allele    GB:AF125672    343    7431   7431  A>G
             source         isSNP  SNP00101013
             consequence    GB:AF125672.1       344    Silent        2425-2425     S
   Allele    GB:AF125672    343    7751   7751  A>G
             source         isSNP  SNP00101014
             consequence    GB:AF125672.1       344    3'
   Allele    GB:AF125672    343    8597   8597  A>G
             source         isSNP  SNP00062569
             consequence    GB:AF125672.1       344    3'
   Allele    GB:AF125672    343    8602   8602  A>G
             source         isSNP  SNP00012487
             consequence    GB:AF125672.1       344    3'
   GIF  NCOR2-cdna-fwd.gif


NOG
Full name : NOG
Link : NOG_link_genomic
   Subsequence       GB:AC005553  1       179651        #345
   Subsequence       NOG_cds.1    146202        145504        #346
   Subsequence       NOG_mrna_build.1   147012        145466        #347
   CDS  NOG_cds.1    699 bp       1 exon        #346
       exon   146202        145504
   mRNA        NOG_mrna_build.1    1547 bp       1 exon        #347
       exon   147012        145466
   Allele      GB:AC005553  345    145585        145585        A>G
               source       wetSNP        GB:AC005553.v145585.G>A
               consequence  NOG_cds.1    346    Silent        206-206       R
   GIF  NOG-genomic-rev.gif


NOTCH3
Link : NOTCH3_link_cdna
   Subsequence       GB:NOTCH3     1      8091   #348
   CDS  GB:NOTCH3.1  6966 bp              #349
       ORF    79      7044
   Allele      GB:NOTCH3     348    1218   1218   A>G
               source        isSNP  SNP00116668
               consequence   GB:NOTCH3.1  349    Silent        380-380       P
   Allele      GB:NOTCH3     348    1565   1565   A>G
               source        isSNP  SNP00116669
               consequence   GB:NOTCH3.1  349    Missense      496-496       P>L
   Allele      GB:NOTCH3     348    2616   2616   A>G
               source        isSNP  SNP00116670
               consequence   GB:NOTCH3.1  349    Silent        846-846       C
   Allele      GB:NOTCH3     348    4520   4520   A>G
               source        isSNP  SNP00116671
               consequence   GB:NOTCH3.1  349    Missense      1481-1481     D>G
   Allele      GB:NOTCH3     348    5740   5740   A>G
               source        isSNP  SNP00054178
               consequence   GB:NOTCH3.1  349    Missense      1888-1888     F>L
```

TABLE 1 (Cont.)

```
  Allele     GB:NOTCH3    348    6355   6355   A>G
             source       isSNP  SNP00037780
             consequence  GB:NOTCH3.1  349    Missense     2093-2093    A>T
  Allele     GB:NOTCH3    348    6516   6516   A>G
             source       isSNP  SNP00054179
             consequence  GB:NOTCH3.1  349    Silent       2146-2146    A
  Allele     GB:NOTCH3    348    6746   6746   A>G
             source       isSNP  SNP00048081
             consequence  GB:NOTCH3.1  349    Missense     2223-2223    V>A
  Allele     GB:NOTCH3    348    7733   7733   A>G
             source       isSNP  SNP00037781
             consequence  GB:NOTCH3.1  349    3'
  Allele     GB:NOTCH3    348    7881   7881   A>G
             source       isSNP  SNP00062225
             consequence  GB:NOTCH3.1  349    3'
  Allele     GB:NOTCH3    348    7914   7914   A>G
             source       isSNP  SNP00066446
             consequence  GB:NOTCH3.1  349    3'
  Allele     GB:NOTCH3    348    8023   8023   A>G
             source       isSNP  SNP00066447
             consequence  GB:NOTCH3.1  349    3'
  GIF NOTCH3-cdna-fwd.gif
Link : NOTCH3_link_genomic
  Subsequence       NOTCH3_cds.1       40735  3819   #350
  Subsequence       GB:AC004663_1      1      41150  #351
  CDS NOTCH3_cds.1        6846 bp      32 exons      #350
      exon    40733 40657
      exon    35676 35534
      exon    35455 35117
      exon    35024 34902
      exon    34814 34581
      exon    32585 32430
      exon    32331 32146
      exon    31505 31392
      exon    31151 31038
      exon    30495 30262
      exon    30145 30035
      exon    28836 28644
      exon    28565 28414
      exon    28176 28063
      exon    27607 27452
      exon    24958 24733
      exon    24319 24118
      exon    23985 23838
      exon    23413 23229
      exon    22653 22521
      exon    22439 22182
      exon    22098 21980
      exon    21247 20682
      exon    17557 17225
      exon    13982 13828
      exon    13710 13488
      exon    13327 13243
      exon    10568 10406
      exon    9248  8944
```

TABLE 1 (Cont.)

	exon	8672	8525					
	exon	5719	5622					
	exon	4871	3819					
Allele	GB:AC004663_1	351	3796	3796	A>T			
	source	wetSNP	GB:AC004663_1.v3796.A>T					
	consequence	NOTCH3_cds.1		350	3′			
Allele	GB:AC004663_1	351	4117	4117	A>G			
	source	isSNP SNP00048081						
	consequence	NOTCH3_cds.1		350	Missense	2183-2183	A>V	
Allele	GB:AC004663_1	351	4347	4347	A>G			
	source	isSNP SNP00054179						
	consequence	NOTCH3_cds.1		350	Silent	2106-2106	A	
Allele	GB:AC004663_1	351	4508	4508	A>G			
	source	isSNP SNP00037780						
	consequence	NOTCH3_cds.1		350	Missense	2053-2053	A>T	
Allele	GB:AC004663_1	351	5727	5727	A>G			
	source	wetSNP	GB:AC004663_1.v5727.A>G					
	consequence	NOTCH3_cds.1		350	Intron			
Allele	GB:AC004663_1	351	5943	5943	A>G			
	source	dbSNP gnl|dbSNP|ss730238_allele						
	consequence	NOTCH3_cds.1		350	Intron			
Allele	GB:AC004663_1	351	17519	17519	A>G			
	source	isSNP SNP00116671						
	consequence	NOTCH3_cds.1		350	Missense	1441-1441	D>G	
Allele	GB:AC004663_1	351	18749	18749	A>G			
	source	dbSNP gnl|dbSNP|ss680542_allele						
	source	dbSNP gnl|dbSNP|ss1143619_allele						
	source	dbSNP gnl|dbSNP|ss372819_allele						
	consequence	NOTCH3_cds.1		350	Intron			
Allele	GB:AC004663_1	351	22353	22353	A>G			
	source	wetSNP	GB:AC004663_1.v22353.C>T					
	consequence	NOTCH3_cds.1		350	Missense	1143-1143	V>M	
Allele	GB:AC004663_1	351	23922	23922	C>G			
	source	wetSNP	GB:AC004663_1.v23922.C>G					
	consequence	NOTCH3_cds.1		350	Missense	980-980	A>P	
Allele	GB:AC004663_1	351	24045	24045	A>G			
	source	wetSNP	GB:AC004663_1.v24045.T>C					
	consequence	NOTCH3_cds.1		350	Intron			
Allele	GB:AC004663_1	351	27480	27480	A>G			
	source	isSNP SNP00116670						
	consequence	NOTCH3_cds.1		350	Silent	806-806	C	
Allele	GB:AC004663_1	351	28173	28173	A>G			
	source	wetSNP	GB:AC004663_1.v28173.C>T					
	consequence	NOTCH3_cds.1		350	Missense	727-727	R>H	
Allele	GB:AC004663_1	351	28749	28749	A>G			
	source	wetSNP	GB:AC004663_1.v28749.C>T					
	consequence	NOTCH3_cds.1		350	Missense	640-640	R>H	
Allele	GB:AC004663_1	351	29997	29997	C>G			
	source	wetSNP	GB:AC004663_1.v29997.G>C					
	consequence	NOTCH3_cds.1		350	Intron			
Allele	GB:AC004663_1	351	32482	32482	A>G			
	source	isSNP SNP00116668						
	consequence	NOTCH3_cds.1		350	Silent	340-340	P	

GIF NOTCH3-genomic-rev.gif

TABLE 1 (Cont.)

```
NPR2
Full name : Atrionatriuretic Peptide Receptor Type B
Link : NPR2_link_cdna
  Subsequence      GB:HUMGUANCYC       1      4081   #352
  CDS GB:HUMGUANCYC.2     3144 bp             #353
      ORF    651    3794
  Allele      GB:HUMGUANCYC      352    2222   2222   A>G
              source       isSNP  SNP00028343
              consequence  GB:HUMGUANCYC.2    353    Silent       524-524      Y
  GIF NPR2-cdna-fwd.gif



OGN
Full name : osteoglycin
Link : OGN_link_cdna
  Subsequence      GB:HSM801395        1      2101   #354
  CDS GB:HSM801395.1      441 bp              #355
      ORF    1      441
  Allele      GB:HSM801395       354    64     64     A>G
              source       isSNP  SNP00100803
              consequence  GB:HSM801395.1     355    Missense     22-22  L>F
  Allele      GB:HSM801395       354    909    909    A>G
              source       isSNP  SNP00011097
              consequence  GB:HSM801395.1     355    3'
  GIF OGN-cdna-fwd.gif
Link : OGN_link_genomic
  Subsequence      OGN_cds.2     48897  32003  #356
  Subsequence      GB:AL354924_2       1      192427      #357
  Subsequence      OGN_mrna_build.2    50083  30350  #358
  mRNA        OGN_mrna_build.2   2726 bp       7 exons      #358
      exon   50083  49983
      exon   48969  48721
      exon   46672  46579
      exon   38619  38461
      exon   35431  35229
      exon   32679  32584
      exon   32173  30350
  CDS OGN_cds.2      900 bp       6 exons      #356
      exon   48897  48721
      exon   46672  46579
      exon   38619  38461
      exon   35431  35229
      exon   32679  32584
      exon   32173  32003
  Allele      GB:AL354924_2      357    31535  31535  A>G
              source       isSNP  SNP00011097
              consequence  OGN_cds.2     356    3'
  Allele      GB:AL354924_2      357    35339  35339  A>G
              source       isSNP  SNP00100803
              consequence  OGN_cds.2     356    Missense     175-175      L>F
  GIF OGN-genomic-rev.gif
```

TABLE 1 (Cont.)

```
OMD
Full name : osteomodulin
Link : OMD_link_cdna
  Subsequence      GB:OMD       1      2263   #359
  CDS GB:OMD.1     1266 bp             #360
      ORF    101    1366
  Allele      GB:OMD        359    159    159    C>G
              source        isSNP  SNP00023658
              consequence   GB:OMD.1      360    Missense     20-20  C>S
  Allele      GB:OMD        359    762    762    A>G
              source        isSNP  SNP00023659
              consequence   GB:OMD.1      360    Missense     221-221      S>N
  Allele      GB:OMD        359    1969   1969   A>G
              source        isSNP  SNP00023660
              consequence   GB:OMD.1      360    3'
  Allele      GB:OMD        359    2071   2071   G>T
              source        isSNP  SNP00106046
              consequence   GB:OMD.1      360    3'
  GIF OMD-cdna-fwd.gif
Link : FL_1258977_link_genomic
  Subsequence      GB:AB009589  1      12414  #361
  Subsequence      GB:AB009589_1258977CD1    8540   10946  #362
  Subsequence      FL_1258977_mrna_build.1   1685   11855  #363
  mRNA        FL_1258977_mrna_build.1  2396 bp     3 exons      #363
       exon   1685   1892
       exon   8524   9479
       exon   10624  11855
  CDS GB:AB009589_1258977CD1    1263 bp      2 exons      #362
       exon   8540   9479
       exon   10624  10946
  Allele      GB:AB009589   361    8598   8598   C>G
              source        isSNP  SNP00023658
              consequence   GB:AB009589_1258977CD1    362    Missense     20-20  C>S
  Allele      GB:AB009589   361    9201   9201   A>G
              source        isSNP  SNP00023659
              consequence   GB:AB009589_1258977CD1    362    Missense     221-221
      S>N
  Allele      GB:AB009589   361    10042  10042  A>G
              source        dbSNP  gnl|dbSNP|ss312223_allele
              consequence   GB:AB009589_1258977CD1    362    Intron
  Allele      GB:AB009589   361    10596  10596  A>G
              source        wetSNP        GB:AB009589.v10596.A>G
              consequence   GB:AB009589_1258977CD1    362    Intron
  Allele      GB:AB009589   361    11552  11552  A>G
              source        isSNP  SNP00023660
              consequence   GB:AB009589_1258977CD1    362    3'
  Allele      GB:AB009589   361    11654  11654  G>T
              source        isSNP  SNP00106046
              consequence   GB:AB009589_1258977CD1    362    3'
  GIF OMD-genomic-fwd.gif


PDCD6IP
Full name : programmed cell death 6-interacting protein
```

TABLE 1 (Cont.)

```
Link : PDCD6IP_link_cdna
  Subsequence       GB:AF151793  1      3221    #364
  CDS  GB:AF151793.1        2607 bp             #365
       ORF   127    2733
  Allele      GB:AF151793  364   1051   1051   A>G
              source       isSNP SNP00029958
              consequence  GB:AF151793.1        365    Missense     309-309      T>A
  Allele      GB:AF151793  364   1258   1258   A>G
              source       isSNP SNP00108790
              consequence  GB:AF151793.1        365    Missense     378-378      V>I
  Allele      GB:AF151793  364   1298   1298   G>T
              source       isSNP SNP00108791
              consequence  GB:AF151793.1        365    Missense     391-391      L>W
  Allele      GB:AF151793  364   1695   1695   A>G
              source       isSNP SNP00093444
              consequence  GB:AF151793.1        365    Silent       523-523      L
  Allele      GB:AF151793  364   2230   2230   A>G
              source       isSNP SNP00121559
              consequence  GB:AF151793.1        365    Missense     702-702      R>G
  Allele      GB:AF151793  364   2315   2315   A>G
              source       isSNP SNP00006604
              consequence  GB:AF151793.1        365    Missense     730-730      L>S
  Allele      GB:AF151793  364   2386   2386   A>G
              source       isSNP SNP00029960
              consequence  GB:AF151793.1        365    Missense     754-754      P>S
  Allele      GB:AF151793  364   2421   2421   A>G
              source       isSNP SNP00121560
              consequence  GB:AF151793.1        365    Silent       765-765      A
  GIF  PDCD6IP-cdna-fwd.gif


PDNP1
Full name : phosphodiesterase I (nucleotide pyrophosphatase I (homologous
to mouse Ly-41 antigen))
Link : PDNP1_link_cdna
  Subsequence       EM:HSAUTOTAX        1      3231    #366
  CDS  EM:HSAUTOTAX.2       2748 bp            #367
       ORF   50     2797
  Allele      EM:HSAUTOTAX        366    342    342    A>G
              source       isSNP SNP00025434
              consequence  EM:HSAUTOTAX.2       367    Missense     98-98  A>V
  Allele      EM:HSAUTOTAX        366    696    696    A>G
              source       isSNP SNP00075872
              consequence  EM:HSAUTOTAX.2       367    Missense     216-216      T>I
  Allele      EM:HSAUTOTAX        366    1682   1682   A>G
              source       isSNP SNP00025435
              consequence  EM:HSAUTOTAX.2       367    Missense     545-545      P>S
  Allele      EM:HSAUTOTAX        366    1789   1789   A>G
              source       isSNP SNP00004604
              consequence  EM:HSAUTOTAX.2       367    Silent       580-580      H
  Allele      EM:HSAUTOTAX        366    2398   2398   G>T
              source       isSNP SNP00122211
              consequence  EM:HSAUTOTAX.2       367    Silent       783-783      V
  Allele      EM:HSAUTOTAX        366    2539   2539   A>G
```

TABLE 1 (Cont.)

```
          source        isSNP  SNP00004605
          consequence  EM:HSAUTOTAX.2      367    Silent        830-830      F
  Allele  EM:HSAUTOTAX         366    2681   2681   G>T
          source        isSNP  SNP00059344
          consequence  EM:HSAUTOTAX.2      367    Silent        878-878      R
  GIF  PDNP1-cdna-fwd.gif
Link  :  PDNP1_link_genomic
  Subsequence       IN:98092911313498  4217   4948   #368
  Subsequence       IN:98061109562226435      5050   5980   #369
  Subsequence       IN:98092910591328158      3611   4115   #370
  Subsequence       IN:98092911013628201      100    699    #371
  Subsequence       IN:98092911024828217      2027   2526   #372
  Subsequence       IN:98092911044928261      3068   3509   #373
  Subsequence       IN:98092911065328292      801    1418   #374
  Subsequence       IN:98092913141116289      6183   6572   #375
  Subsequence       IN:98111010592914993      1520   1926   #376
  Subsequence       IN:98111011021915028      2628   2967   #377
  Allele    IN:98092910591328158      370    232    232    A>G
            source        isSNP  SNP00025435
  Allele    IN:98092913141116289      375    189    189    G>T
            source        isSNP  SNP00059344



PLA2G2A
Full name : phospholipase A2, group IIA
Link  :  PLA2G2A_link_cdna
  Subsequence       GB:HUMRASFAB        1      854    #378
  CDS  GB:HUMRASFAB.1      435  bp             #379
     ORF    136    570
  Allele     GB:HUMRASFAB        378    267    267    A>G
             source        isSNP  SNP00010003
             consequence  GB:HUMRASFAB.1      379    Silent        44-44  Y
  Allele     GB:HUMRASFAB        378    800    800    A>G
             source        isSNP  SNP00021612
             consequence  GB:HUMRASFAB.1      379    3'
  GIF  PLA2G2A-cdna-fwd.gif
Link  :  PLA2G2A_link_genomic
  Subsequence       PLA2G2A_cds.1       51704  48629  #380
  Subsequence       PLA2G2A_mrna_build.1       52537  48418  #381
  Subsequence       GB:AL358253_1       1      180550        #382
  Subsequence       LG:474322.13_mrna_build.1       52786  48418  #383
  Subsequence       PLA2G2A_cds.2       51704  50985  #384
  mRNA      LG:474322.13_mrna_build.1      1028  bp       5  exons       #383
       exon   52786  52511
       exon   51810  51665
       exon   51455  51311
       exon   51052  50946
       exon   48771  48418
  CDS  PLA2G2A_cds.1      435  bp       4  exons       #380
       exon   51704  51665
       exon   51455  51311
       exon   51052  50946
       exon   48771  48629
  CDS  PLA2G2A_cds.2      108  bp       2  exons       #384
```

TABLE 1 (Cont.)

```
      exon   51704  51665
      exon   51052  50985
  mRNA       PLA2G2A_mrna_build.1     779 bp       5 exons       #381
      exon   52537  52511
      exon   51810  51665
      exon   51455  51311
      exon   51052  50946
      exon   48771  48418
  Allele     GB:AL358253_1       382    51364  51364  A>G
             source       isSNP  SNP00010003
             consequence  PLA2G2A_cds.1      380    Silent       44-44  Y
             consequence  PLA2G2A_cds.2      384    Intron
  Allele     GB:AL358253_1       382    52584  52584  C>G
             source       isSNP  SNP00021611
             consequence  PLA2G2A_cds.1      380    5'
             consequence  PLA2G2A_cds.2      384    5'
  GIF  PLA2G2A-genomic-rev.gif


PPP1R5
Full name : protein phosphatase 1, regulatory (inhibitor) subunit 5
Link : PPP1R5_link_cdna
  Subsequence      GB:Y18207_1  1      1158   #385
  CDS  GB:Y18207_1.1      954 bp              #386
       ORF    92     1045
  Allele     GB:Y18207_1  385    571    571    A>G
             source       isSNP  SNP00041149
             consequence  GB:Y18207_1.1      386    Silent       160-160      E
  Allele     GB:Y18207_1  385    1096   1096   G>T
             source       isSNP  SNP00060710
             consequence  GB:Y18207_1.1      386    3'
  GIF  PPP1R5-cdna-fwd.gif
Link : PPP1R5_link_genomic
  Subsequence      GB:AC020691_2         1        152048        #387
  Subsequence      PPP1R5_mrna_build.1          103997        107245       #388
  Subsequence      PPP1R5_cds.1          106194        107132       #389
  CDS  PPP1R5_cds.1        939 bp        1 exon        #389
       exon   106194       107132
  mRNA       PPP1R5_mrna_build.1        1160 bp       2 exons       #388
       exon   103997       104103
       exon   106193       107245
  Allele     GB:AC020691_2        387    106523        106523        G>T
             source       wetSNP        GB:AC020691_2.v106523.T>G
             consequence  PPP1R5_cds.1         389    Missense      110-110      D>E
  Allele     GB:AC020691_2        387    106658        106658        A>G
             source       isSNP  SNP00041149
             consequence  PPP1R5_cds.1         389    Silent        155-155      E
  Allele     GB:AC020691_2        387    107183        107183        G>T
             source       isSNP  SNP00060710
             consequence  PPP1R5_cds.1         389    3'
  GIF  PPP1R5-genomic-fwd.gif
```

TABLE 1 (Cont.)

```
PRELP
Full name : proline arginine-rich end leucine-rich repeat protein
Link : PRELP_link_cdna
  Subsequence      GB:HSU29089  1      1560      #390
  CDS  GB:HSU29089.1      1149 bp                #391
      ORF    129    1277
  Allele      GB:HSU29089  390    1170   1170    G>T
              source        isSNP  SNP00001359
              consequence  GB:HSU29089.1         391    Missense     348-348      N>H
  Allele      GB:HSU29089  390    1489   1489    G>T
              source        isSNP  SNP00001361
              consequence  GB:HSU29089.1         391    3'
  GIF  PRELP-cdna-fwd.gif
Link : PRELP_link_genomic
  Subsequence      PRELP_cds.1  82496  86192  #392
  Subsequence      GB:AC022000_1       1      154681       #393
  Subsequence      PRELP_mrna_build.1         75139  86474  #394
  CDS  PRELP_cds.1  1149 bp      2 exons      #392
      exon   82496  83468
      exon   86017  86192
  mRNA        PRELP_mrna_build.1        1559 bp      3 exons       #394
      exon   75139  75250
      exon   82480  83468
      exon   86017  86474
  Allele      GB:AC022000_1       393    86085  86085  G>T
              source        isSNP  SNP00001359
              consequence  PRELP_cds.1  392    Missense     348-348      N>H
  Allele      GB:AC022000_1       393    86404  86404  G>T
              source        isSNP  SNP00001361
              consequence  PRELP_cds.1  392    3'
  GIF  PRELP-genomic-fwd.gif


PRSS11
Full name : serine protease.
Link : FL_1787335_link_cdna
  Subsequence      FN:1787335CB1       1      2054    #395
  CDS  FN:1787335CB1.1     1443  bp              #396
      ORF    49     1491
  Allele      FN:1787335CB1       395    150     150     A>G
              source        isSNP  SNP00068999
              consequence  FN:1787335CB1.1      396    Silent       34-34  A
  Allele      FN:1787335CB1       395    156     156     G>T
              source        isSNP  SNP00117078
              consequence  FN:1787335CB1.1      396    Silent       36-36  G
  Allele      FN:1787335CB1       395    914     914     A>G
              source        isSNP  SNP00120314
              consequence  FN:1787335CB1.1      396    Missense     289-289      Q>R
  Allele      FN:1787335CB1       395    1321    1321    C>G
              source        isSNP  SNP00105589
              consequence  FN:1787335CB1.1      396    Missense     425-425      A>P
  Allele      FN:1787335CB1       395    1521    1521    A>G
              source        isSNP  SNP00105590
              consequence  FN:1787335CB1.1      396    3'
```

TABLE 1 (Cont.)

```
  GIF  PRSS11-cdna-fwd.gif
Link : FL_1787335_link_genomic
  Subsequence      GB:AF157623_1_1787335CD1        17526  70213  #397
  Subsequence      GB:AF157623_1       1       79597  #398
  Subsequence      FL_1787335_mrna_build.1   17478  70761  #399
  CDS  GB:AF157623_1_1787335CD1        1443 bp      9 exons       #397
       exon   17526  17997
       exon   44770  44869
       exon   45290  45494
       exon   62561  62755
       exon   63240  63272
       exon   64526  64640
       exon   65966  66023
       exon   67827  67922
       exon   70045  70213
  mRNA         FL_1787335_mrna_build.1   2039 bp      9 exons       #399
       exon   17478  17997
       exon   44770  44869
       exon   45290  45494
       exon   62561  62755
       exon   63240  63272
       exon   64526  64640
       exon   65966  66023
       exon   67827  67922
       exon   70045  70761
   Allele      GB:AF157623_1       398    17627  17627  A>G
               source        isSNP  SNP00068999
               consequence  GB:AF157623_1_1787335CD1        397    Silent
 34-34  A
   Allele      GB:AF157623_1       398    17633  17633  G>T
               source        isSNP  SNP00117078
               consequence  GB:AF157623_1_1787335CD1        397    Silent
 36-36  G
   Allele      GB:AF157623_1       398    21721  21721  A>G
               source        isSNP  SNP00101582
               consequence  GB:AF157623_1_1787335CD1        397    Intron
   Allele      GB:AF157623_1       398    35790  35790  A>G
               source        isSNP  SNP00049308
               consequence  GB:AF157623_1_1787335CD1        397    Intron
   Allele      GB:AF157623_1       398    44762  44762  G>T
               source        wetSNP       GB:AF157623_1.v44762.G>T
               consequence  GB:AF157623_1_1787335CD1        397    Intron
   Allele      GB:AF157623_1       398    45470  45470  A>G
               source        wetSNP       GB:AF157623_1.v45470.C>T
               consequence  GB:AF157623_1_1787335CD1        397    Silent
 251-251       I
   Allele      GB:AF157623_1       398    45587  45587  A>G
               source        wetSNP       GB:AF157623_1.v45587.C>T
               consequence  GB:AF157623_1_1787335CD1        397    Intron
   Allele      GB:AF157623_1       398    47792  47792  A>G
               source        isSNP  SNP00105588
               consequence  GB:AF157623_1_1787335CD1        397    Intron
   Allele      GB:AF157623_1       398    47834  47834  A>G
               source        isSNP  SNP00120312
               consequence  GB:AF157623_1_1787335CD1        397    Intron
```

TABLE 1 (Cont.)

Allele GB:AF157623_1 398 47913 47913 A>G
source isSNP SNP00120313
consequence GB:AF157623_1_1787335CD1 397 Intron
Allele GB:AF157623_1 398 62541 62541 A>G
source wetSNP GB:AF157623_1.v62541.G>A
consequence GB:AF157623_1_1787335CD1 397 Intron
Allele GB:AF157623_1 398 62545 62545 A>G
source wetSNP GB:AF157623_1.v62545.G>A
consequence GB:AF157623_1_1787335CD1 397 Intron
Allele GB:AF157623_1 398 62649 62649 A>G
source isSNP SNP00120314
consequence GB:AF157623_1_1787335CD1 397 Missense
289-289 Q>R
Allele GB:AF157623_1 398 63355 63360 TGTTTT>TT
source wetSNP GB:AF157623_1.v63355.TGTTTT>TT
consequence GB:AF157623_1_1787335CD1 397 Intron
Allele GB:AF157623_1 398 70243 70243 A>G
source isSNP SNP00105590
consequence GB:AF157623_1_1787335CD1 397 3′
GIF PRSS11-genomic-fwd.gif

PTGS2
Full name : Prostaglandin-endoperoxide Synthase 2
Link : PTGS2_link_cdna
Subsequence EM:HSCYCLOX 1 3387 #400
Allele EM:HSCYCLOX 400 403 403 C>G
source isSNP SNP00046167
Allele EM:HSCYCLOX 400 880 880 G>T
source isSNP SNP00076329
Allele EM:HSCYCLOX 400 2033 2033 A>G
source isSNP SNP00076330
Allele EM:HSCYCLOX 400 2300 2300 A>G
source isSNP SNP00046168
Allele EM:HSCYCLOX 400 2983 2983 A>G
source isSNP SNP00046169
Link : PTGS2_link_genomic
Subsequence GB:HUMPTGS2 101 11097 #401
Subsequence PTGS2_cds.1 1925 8146 #402
Subsequence PTGS2_mrna_build.1 1828 9607 #403
CDS PTGS2_cds.1 1815 bp 10 exons #402
exon 1925 1976
exon 2777 2893
exon 3014 3157
exon 3811 3954
exon 4670 4851
exon 5584 5667
exon 5787 6033
exon 6315 6601
exon 7103 7250
exon 7737 8146
mRNA PTGS2_mrna_build.1 3373 bp 10 exons #403
exon 1828 1976
exon 2777 2893

TABLE 1 (Cont.)

	exon	3014	3157				
	exon	3811	3954				
	exon	4670	4851				
	exon	5584	5667				
	exon	5787	6033				
	exon	6315	6601				
	exon	7103	7250				
	exon	7737	9607				
Allele	GB:HUMPTGS2	401	3050	3050	C>G		
	source	wetSNP	GB:HUMPTGS2.v3050.G>C				
	consequence	PTGS2_cds.1	402	Silent	102-102	V	
Allele	GB:HUMPTGS2	401	3090	3090	A>G		
	source	wetSNP	GB:HUMPTGS2.v3090.C>T				
	consequence	PTGS2_cds.1	402	Intron			
Allele	GB:HUMPTGS2	401	3174	3174	C>G		
	source	wetSNP	GB:HUMPTGS2.v3174.G>C				
	consequence	PTGS2_cds.1	402	Intron			
Allele	GB:HUMPTGS2	401	3793	3793	A>G		
	source	wetSNP	GB:HUMPTGS2.v3793.C>T				
	consequence	PTGS2_cds.1	402	Silent	132-132	S	
Allele	GB:HUMPTGS2	401	3829	3829	A>G		
	source	wetSNP	GB:HUMPTGS2.v3829.T>C				
	consequence	PTGS2_cds.1	402	Silent	144-144	D	
Allele	GB:HUMPTGS2	401	5605	5605	A>G		
	source	wetSNP	GB:HUMPTGS2.v5605.G>A				
	consequence	PTGS2_cds.1	402	Intron			
Allele	GB:HUMPTGS2	401	5676	5681	TATTTT>TT		
	source	wetSNP	GB:HUMPTGS2.v5676.TATTTT>TT				
	consequence	PTGS2_cds.1	402	Intron			
Allele	GB:HUMPTGS2	401	5746	5746	G>T		
	source	isSNP	SNP00076329				
	consequence	PTGS2_cds.1	402	Stop	261-261		
Allele	GB:HUMPTGS2	401	6249	6249	A>G		
	source	wetSNP	GB:HUMPTGS2.v6249.G>A				
	consequence	PTGS2_cds.1	402	Silent	335-335	V	
Allele	GB:HUMPTGS2	401	6444	6444	A>G		
	source	wetSNP	GB:HUMPTGS2.v6444.G>A				
	consequence	PTGS2_cds.1	402	Silent	400-400	L	
Allele	GB:HUMPTGS2	401	6453	6453	A>G		
	source	wetSNP	GB:HUMPTGS2.v6453.T>C				
	consequence	PTGS2_cds.1	402	Silent	403-403	H	
Allele	GB:HUMPTGS2	401	7581	7581	A>G		
	source	wetSNP	GB:HUMPTGS2.v7581.T>C				
	consequence	PTGS2_cds.1	402	Intron			
Allele	GB:HUMPTGS2	401	7763	7763	A>G		
	source	wetSNP	GB:HUMPTGS2.v7763.T>C				
	consequence	PTGS2_cds.1	402	Missense	511-511	V>A	
Allele	GB:HUMPTGS2	401	7986	7986	G>T		
	source	wetSNP	GB:HUMPTGS2.v7986.C>A				
	consequence	PTGS2_cds.1	402	Silent	585-585	R	
Allele	GB:HUMPTGS2	401	8167	8167	A>G		
	source	isSNP	SNP00076330				
	consequence	PTGS2_cds.1	402	3'			
Allele	GB:HUMPTGS2	401	8434	8434	A>G		
	source	isSNP	SNP00046168				

TABLE 1 (Cont.)

consequence PTGS2_cds.1 402 3'
Allele GB:HUMPTGS2 401 8473 8473 A>G
source isSNP SNP00012871
consequence PTGS2_cds.1 402 3'
Allele GB:HUMPTGS2 401 9102 9102 A>G
source isSNP SNP00046169
consequence PTGS2_cds.1 402 3'
GIF PTGS2-genomic-fwd.gif

PTHLH
Full name : PTHLH
Link : PTHLH_link_genomic
Subsequence PTHLH_cds.1 106964 117899 #404
Subsequence GB:AC008011_6 1 183178 #405
Subsequence PTHLH_mrna_build.1 106942 118367 #406
CDS PTHLH_cds.1 534 bp 3 exons #404
exon 106964 107064
exon 112688 113110
exon 117890 117899
mRNA PTHLH_mrna_build.1 1024 bp 3 exons #406
exon 106942 107064
exon 112688 113110
exon 117890 118367
Allele GB:AC008011_6 405 113450 113450 A>G
source isSNP SNP00043978
consequence PTHLH_cds.1 404 Intron
Allele GB:AC008011_6 405 115075 115075 A>G
source dbSNP gnl|dbSNP|ss1455356_allele
consequence PTHLH_cds.1 404 Intron
Allele GB:AC008011_6 405 115160 115160 A>G
source dbSNP gnl|dbSNP|ss1067559_allele
consequence PTHLH_cds.1 404 Intron
GIF PTHLH-genomic-fwd.gif

PTHR1
Full name : PTHR1
Link : PTHR1_link_cdna
Subsequence GB:HUMPTHR 1 1948 #407
CDS GB:HUMPTHR.1 1782 bp #408
ORF 29 1810
Allele GB:HUMPTHR 407 1417 1417 A>G
source isSNP SNP00007059
consequence GB:HUMPTHR.1 408 Silent 463-463 N
GIF PTHR1-cdna-fwd.gif
Link : PTHR1_link_genomic
Subsequence GB:HSPTHPRH1 1 262 #409
Subsequence GB:HSPTHPRH2 363 769 #410
Subsequence GB:HSPTHPRH3 870 1168 #411
Subsequence GB:HSPTHPRH4 1269 2146 #412
Subsequence GB:HSPTHPRH5 2247 3249 #413
Subsequence GB:HSPTHPRH6 3350 4062 #414

TABLE 1 (Cont.)

```
  Subsequence      GB:HSPTHPRH7       4163   4475   #415
  Subsequence      GB:HSPTHPRH8       4576   4995   #416
  Subsequence      GB:HSPTHPRH9       5096   5696   #417
  Subsequence      PTHR1_cds.1  107    5558   #418
  Subsequence      PTHR1_mrna_build.1         79     5696   #419
  CDS  PTHR1_cds.1  1782 bp      14 exons      #418
       exon   107    181
       exon   456    558
       exon   936    1070
       exon   1436   1546
       exon   1655   1773
       exon   1959   2053
       exon   2351   2546
       exon   2980   3133
       exon   3547   3607
       exon   3938   4004
       exon   4273   4367
       exon   4628   4769
       exon   4851   4892
       exon   5172   5558
  mRNA        PTHR1_mrna_build.1        1948 bp      14 exons      #419
       exon   79     181
       exon   456    558
       exon   936    1070
       exon   1436   1546
       exon   1655   1773
       exon   1959   2053
       exon   2351   2546
       exon   2980   3133
       exon   3547   3607
       exon   3938   4004
       exon   4273   4367
       exon   4628   4769
       exon   4851   4892
       exon   5172   5696
  Allele      GB:HSPTHPRH3        411    104    104    A>G
              source        wetSNP      GB:HSPTHPRH3.v104.G>A
              consequence  PTHR1_cds.1  418    Silent       72-72  A
  Allele      GB:HSPTHPRH8        416    311    311    A>G
              source        wetSNP      GB:HSPTHPRH8.v311.T>C
              consequence  PTHR1_cds.1  418    Silent       463-463      N
  GIF  PTHR1-genomic-fwd.gif


RARA
Full name : retinoic acid receptor, alpha
Link : RARA_link_cdna
  Subsequence      GB:NM_000964       1      2907   #420
  CDS  GB:NM_000964.1      1389 bp             #421
       ORF    103    1491
  Allele      GB:NM_000964        420    2327   2327   A>G
              source        isSNP  SNP00016145
              consequence  GB:NM_000964.1       421    3'
  Allele      GB:NM_000964        420    2439   2439   A>G
```

TABLE 1 (Cont.)

```
          source        isSNP  SNP00049381
          consequence   GB:NM_000964.1      421    3'
  GIF  RARA-cdna-fwd.gif


RIN1
Full name : ras inhibitor
Link : RIN1_link_cdna
  Subsequence      GB:HUMRASINF        1      1285    #422
  Allele      GB:HUMRASINF        422    260     260     A>G
              source        isSNP  SNP00123606
  Allele      GB:HUMRASINF        422    424     424     A>G
              source        isSNP  SNP00123607
  Allele      GB:HUMRASINF        422    722     722     A>G
              source        isSNP  SNP00033587
  Allele      GB:HUMRASINF        422    921     921     A>G
              source        isSNP  SNP00007808


ROR2
Full name : receptor tyrosine kinase-like orphan receptor 2
Link : ROR2_link_cdna
  Subsequence      GB:NM_004560        1      4092    #423
  CDS  GB:NM_004560.1     2832 bp             #424
       ORF    200    3031
  Allele      GB:NM_004560        423    932     932     A>G
              source        isSNP  SNP00098926
              consequence   GB:NM_004560.1      424    Missense     245-245     A>T
  Allele      GB:NM_004560        423    1460    1460    A>G
              source        isSNP  SNP00098927
              consequence   GB:NM_004560.1      424    Missense     421-421     L>F
  Allele      GB:NM_004560        423    1973    1973    A>G
              source        isSNP  SNP00098928
              consequence   GB:NM_004560.1      424    Missense     592-592     F>L
  Allele      GB:NM_004560        423    2287    2287    A>G
              source        isSNP  SNP00028168
              consequence   GB:NM_004560.1      424    Silent       696-696     Y
  Allele      GB:NM_004560        423    2353    2353    A>G
              source        isSNP  SNP00098929
              consequence   GB:NM_004560.1      424    Silent       718-718     P
  Allele      GB:NM_004560        423    2654    2654    A>G
              source        isSNP  SNP00028169
              consequence   GB:NM_004560.1      424    Missense     819-819     V>I
  Allele      GB:NM_004560        423    3743    3743    A>G
              source        isSNP  SNP00028170
              consequence   GB:NM_004560.1      424    3'
  Allele      GB:NM_004560        423    3872    3872    G>T
              source        isSNP  SNP00074568
              consequence   GB:NM_004560.1      424    3'
  Allele      GB:NM_004560        423    3919    3919    G>T
              source        isSNP  SNP00074569
              consequence   GB:NM_004560.1      424    3'
  GIF  ROR2-cdna-fwd.gif
```

TABLE 1 (Cont.)

```
RORA
Full name : RAR-related orphan receptor alpha
Link : RORA_link_genomic
  Subsequence      RORA_cds.1   64220 3076   #425
  Subsequence      RORA_cds.2   64220 3076   #426
  Subsequence      RORA_cds.4   64220 3076   #427
  Subsequence      GB:AC012344_4_000018      1     9454  #428
  Subsequence      GB:AC012344_4_000020      9555  21185 #429
  Subsequence      GB:AC012344_4_000021      21286 34347 #430
  Subsequence      GB:AC012344_4_000019      34448 43824 #431
  Subsequence      GB:AC012344_4_000023      43925 65900 #432
  Subsequence      RORA_mrna_build.1  64309  2885  #433
  Subsequence      RORA_mrna_build.4  64290  2885  #434
  mRNA       RORA_mrna_build.4  1908 bp     11 exons     #434
       exon  64290 64084
       exon  51847 51714
       exon  25290 25205
       exon  19553 19412
       exon  16417 16022
       exon  10425 10304
       exon  9288  9156
       exon  8488  8381
       exon  6690  6580
       exon  5625  5513
       exon  3240  2885
  CDS  RORA_cds.1   1671 bp     12 exons     #425
       exon  64220 64084
       exon  43229 43148
       exon  41851 41776
       exon  25290 25205
       exon  19553 19412
       exon  16417 16022
       exon  10425 10304
       exon  9288  9156
       exon  8488  8381
       exon  6690  6580
       exon  5625  5513
       exon  3240  3076
  CDS  RORA_cds.2   1275 bp     11 exons     #426
       exon  64220 64084
       exon  43229 43148
       exon  41851 41776
       exon  25290 25205
       exon  19553 19412
       exon  10425 10304
       exon  9288  9156
       exon  8488  8381
       exon  6690  6580
       exon  5625  5513
       exon  3240  3076
  mRNA       RORA_mrna_build.1  1951 bp     12 exons     #433
       exon  64309 64084
       exon  43229 43148
```

TABLE 1 (Cont.)

```
    exon  41851 41776
    exon  25290 25205
    exon  19553 19412
    exon  16417 16022
    exon  10425 10304
    exon  9288  9156
    exon  8488  8381
    exon  6690  6580
    exon  5625  5513
    exon  3240  2885
CDS RORA_cds.4   1647 bp     11 exons    #427
    exon  64220 64084
    exon  51847 51714
    exon  25290 25205
    exon  19553 19412
    exon  16417 16022
    exon  10425 10304
    exon  9288  9156
    exon  8488  8381
    exon  6690  6580
    exon  5625  5513
    exon  3240  3076
Allele      GB:AC012344_4_000020      429    11153 11153 A>G
            source       dbSNP gnl|dbSNP|ss380580_allele
            consequence  RORA_cds.1   425    Intron
            consequence  RORA_cds.2   426    Intron
            consequence  RORA_cds.4   427    Intron
Allele      GB:AC012344_4_000020      429    11182 11182 A>G
            source       dbSNP gnl|dbSNP|ss380580_allele
            consequence  RORA_cds.1   425    Intron
            consequence  RORA_cds.2   426    Intron
            consequence  RORA_cds.4   427    Intron
Allele      GB:AC012344_4_000020      429    11183 11183 A>T
            source       dbSNP gnl|dbSNP|ss507731_allele
            consequence  RORA_cds.1   425    Intron
            consequence  RORA_cds.2   426    Intron
            consequence  RORA_cds.4   427    Intron
Allele      GB:AC012344_4_000020      429    11254 11254 A>G
            source       dbSNP gnl|dbSNP|ss380580_allele
            consequence  RORA_cds.1   425    Intron
            consequence  RORA_cds.2   426    Intron
            consequence  RORA_cds.4   427    Intron
Allele      GB:AC012344_4_000020      429    11255 11255 A>T
            source       dbSNP gnl|dbSNP|ss507731_allele
            consequence  RORA_cds.1   425    Intron
            consequence  RORA_cds.2   426    Intron
            consequence  RORA_cds.4   427    Intron
Allele      GB:AC012344_4_000020      429    11264 11264 A>G
            source       dbSNP gnl|dbSNP|ss380580_allele
            consequence  RORA_cds.1   425    Intron
            consequence  RORA_cds.2   426    Intron
            consequence  RORA_cds.4   427    Intron
Allele      GB:AC012344_4_000020      429    11265 11265 A>T
            source       dbSNP gnl|dbSNP|ss507731_allele
            consequence  RORA_cds.1   425    Intron
```

TABLE 1 (Cont.)

```
          consequence RORA_cds.2   426    Intron
          consequence RORA_cds.4   427    Intron
  Allele  GB:AC012344_4_000020     429    11320 11320 A>G
          source      dbSNP gnl|dbSNP|ss380580_allele
          consequence RORA_cds.1   425    Intron
          consequence RORA_cds.2   426    Intron
          consequence RORA_cds.4   427    Intron
  GIF RORA-genomic-rev.gif



SCRG1
Full name : scrapie responsive protein
Link : SCRG1_link_genomic
  Subsequence     SCRG1_cds.1  30577 33650 #435
  Subsequence     GB:AC009588_4      1     164772     #436
  Subsequence     SCRG1_mrna_build.1       30561 33845 #437
  CDS SCRG1_cds.1 297 bp      2 exons      #435
      exon  30577 30818
      exon  33596 33650
  mRNA        SCRG1_mrna_build.1      508 bp      2 exons      #437
      exon  30561 30818
      exon  33596 33845
  GIF SCRG1-genomic-fwd.gif



SCYA20
Full name : small inducible cytokine subfamily A member 20
Link : SCYA20_link_cdna
  Subsequence     GB:HSU64197  1      821     #438
  CDS GB:HSU64197.1       288 bp              #439
      ORF   43    330
  Allele      GB:HSU64197  438   341   341    A>G
              source       isSNP SNP00037526
              consequence  GB:HSU64197.1      439   3'
  Allele      GB:HSU64197  438   728   728    A>G
              source       isSNP SNP00037527
              consequence  GB:HSU64197.1      439   3'
  GIF SCYA20-cdna-fwd.gif
Link : SCYA20_link_genomic
  Subsequence     SCYA20_cds.1        73925 77096 #440
  Subsequence     GB:AC027560_2       1     129588      #441
  Subsequence     SCYA20_mrna_build.1       73883 77577 #442
  CDS SCYA20_cds.1        288 bp      4 exons     #440
      exon  73925 74000
      exon  75470 75581
      exon  76320 76397
      exon  77075 77096
  mRNA        SCYA20_mrna_build.1     811 bp      4 exons      #442
      exon  73883 74000
      exon  75470 75581
      exon  76320 76397
      exon  77075 77577
  Allele      GB:AC027560_2       441   77107 77107 A>G
```

TABLE 1 (Cont.)

source isSNP SNP00037526
consequence SCYA20_cds.1 440 3'
Allele GB:AC027560_2 441 77493 77493 A>G
source isSNP SNP00037527
consequence SCYA20_cds.1 440 3'
GIF SCYA20-genomic-fwd.gif

SDC2
Full name : syndecan 2
Link : SDC2_link_cdna
Subsequence GB:HUMHSPGC 1 3414 #443
CDS GB:HUMHSPGC.2 1194 bp #444
ORF 1 1194
Allele GB:HUMHSPGC 443 435 435 A>G
source isSNP SNP00116695
consequence GB:HUMHSPGC.2 444 Silent 145-145 P
Allele GB:HUMHSPGC 443 463 463 C>G
source isSNP SNP00050825
consequence GB:HUMHSPGC.2 444 Missense 155-155 L>V
Allele GB:HUMHSPGC 443 741 741 A>G
source isSNP SNP00033651
consequence GB:HUMHSPGC.2 444 Silent 247-247 Y
Allele GB:HUMHSPGC 443 1041 1041 G>T
source isSNP SNP00099428
consequence GB:HUMHSPGC.2 444 Silent 347-347 A
GIF SDC2-cdna-fwd.gif

SDC4
Full name : syndecan 4
Link : FL_1394592_link_cdna
Subsequence FN:1394592CB1 1 2112 #445
CDS FN:1394592CB1.1 594 bp #446
ORF 23 616
CDS GB:HS453C12_1394592CD1 594 bp #272
ORF 87967 88026
ORF 100431 100569
ORF 103282 103328
ORF 105787 105985
ORF 108936 109084
mRNA FL_1394592_mrna_build.1 2110 bp #274
ORF 87945 88026
ORF 100431 100569
ORF 103282 103328
ORF 105787 105985
ORF 108936 110578
Allele FN:1394592CB1 445 653 653 C>G
source isSNP SNP00124074
consequence FN:1394592CB1.1 446 3'
Allele FN:1394592CB1 445 749 749 A>G
source isSNP SNP00124075
consequence FN:1394592CB1.1 446 3'

TABLE 1 (Cont.)

```
  Allele      FN:1394592CB1      445    856    856    A>G
              source        isSNP  SNP00053065
              consequence  FN:1394592CB1.1    446    3'
  Allele      FN:1394592CB1      445    884    884    A>G
              source        isSNP  SNP00066145
              consequence  FN:1394592CB1.1    446    3'
  Allele      FN:1394592CB1      445    1048   1048   A>G
              source        isSNP  SNP00066146
              consequence  FN:1394592CB1.1    446    3'
  Allele      FN:1394592CB1      445    1214   1214   A>G
              source        isSNP  SNP00029910
              consequence  FN:1394592CB1.1    446    3'
  GIF  SDC4-cdna-fwd.gif
Link : FL_1250708_link_genomic
  Subsequence      GB:HS453C12  1      147620        #271
  Subsequence      GB:HS453C12_1394592CD1    87967  109084        #272
  Subsequence      GB:HS453C12_2027624CD1    20194  10528  #273
  Subsequence      FL_1394592_mrna_build.1   87945  110578        #274
  Subsequence      FL_2027624_mrna_build.1   20197  6152   #275
  Subsequence      OA21_cds.1   20194  17050  #276
  CDS  GB:HS453C12_1394592CD1    594 bp       5 exons      #272
       exon   87967  88026
       exon   100431        100569
       exon   103282        103328
       exon   105787        105985
       exon   108936        109084
  mRNA        FL_1394592_mrna_build.1  2110 bp      5 exons      #274
       exon   87945  88026
       exon   100431        100569
       exon   103282        103328
       exon   105787        105985
       exon   108936        110578
  Allele      GB:HS453C12   271    90320  90320  A>G
              source        isSNP  SNP00026142
              consequence   GB:HS453C12_1394592CD1    272    Intron
  Allele      GB:HS453C12   271    90420  90420  C>G
              source        isSNP  SNP00026143
              consequence   GB:HS453C12_1394592CD1    272    Intron
  Allele      GB:HS453C12   271    96768  96768  A>G
              source        dbSNP  gnl|dbSNP|ss736312_allele
              consequence   GB:HS453C12_1394592CD1    272    Intron
  Allele      GB:HS453C12   271    109121        109121       C>G
              source        isSNP  SNP00124074
              consequence   GB:HS453C12_1394592CD1    272    3'
  Allele      GB:HS453C12   271    109217        109217       A>G
              source        isSNP  SNP00124075
              consequence   GB:HS453C12_1394592CD1    272    3'
  Allele      GB:HS453C12   271    109324        109324       A>G
              source        isSNP  SNP00053065
              consequence   GB:HS453C12_1394592CD1    272    3'
  Allele      GB:HS453C12   271    109352        109352       A>G
              source        isSNP  SNP00066145
              consequence   GB:HS453C12_1394592CD1    272    3'
  Allele      GB:HS453C12   271    109516        109516       A>G
              source        isSNP  SNP00066146
```

TABLE 1 (Cont.)

```
            consequence  GB:HS453C12_1394592CD1    272    3'
      Allele      GB:HS453C12  271    109682        109682       A>G
            source       isSNP  SNP00029910
            consequence  GB:HS453C12_1394592CD1    272    3'
      GIF  SDC4-genomic-fwd.gif


SEDL
Full name : sedlin
Link : SEDL_link_cdna
  Subsequence       GB:NM_014563_1      1      2816   #447
  CDS  GB:NM_014563_1.1    423 bp              #448
       ORF    230    652
  Allele      GB:NM_014563_1      447    991    991    G>T
              source       dbSNP  gnl|dbSNP|ss380525_allele
              source       dbSNP  gnl|dbSNP|ss531221_allele
              consequence  GB:NM_014563_1.1    448    3'
  Allele      GB:NM_014563_1      447    2026   2026   A>G
              source       dbSNP  gnl|dbSNP|ss637643_allele
              source       dbSNP  gnl|dbSNP|ss869682_allele
              source       dbSNP  gnl|dbSNP|ss1272499_allele
              source       dbSNP  gnl|dbSNP|ss232503_allele
              source       dbSNP  gnl|dbSNP|ss459122_allele
              consequence  GB:NM_014563_1.1    448    3'
  Allele      GB:NM_014563_1      447    2391   2391   C>G
              source       isSNP  SNP00010387
              consequence  GB:NM_014563_1.1    448    3'
  GIF  SEDL-cdna-fwd.gif


SKI
Full name : v-ski avian sarcoma viral oncogene homolog
Link : SKI_link_cdna
  Subsequence       GB:NM_003036        1      3511   #449
  CDS  GB:NM_003036.1      2187 bp             #450
       ORF    73     2259
  Allele      GB:NM_003036        449    528    528    A>G
              source       isSNP  SNP00068450
              consequence  GB:NM_003036.1      450    Silent       152-152      R
  Allele      GB:NM_003036        449    1146   1146   A>G
              source       isSNP  SNP00068451
              consequence  GB:NM_003036.1      450    Silent       358-358      T
  Allele      GB:NM_003036        449    3482   3482   C>G
              source       isSNP  SNP00068452
              consequence  GB:NM_003036.1      450    3'
  GIF  SKI-cdna-fwd.gif


SOD2
Full name : superoxide dismutase 2, mitochondrial
Link : SOD2_link_cdna
  Subsequence       EM:HSSOD      1      1026   #451
```

TABLE 1 (Cont.)

```
  Allele     EM:HSSOD      451    243    243    A>G
             source        isSNP  SNP00021476
Link : SOD2_link_genomic
  Subsequence        EM:S77127    101    12957  #452
  Subsequence        SOD2_link_cds.1     957    11597  #453
  Subsequence        SOD2_mrna_build.1  953    11950  #454
  mRNA       SOD2_mrna_build.1  1026 bp      5 exons      #454
      exon   953    979
      exon   1260   1462
      exon   5859   5975
      exon   9061   9240
      exon   11452  11950
  CDS  SOD2_link_cds.1    669 bp       5 exons      #453
      exon   957    979
      exon   1260   1462
      exon   5859   5975
      exon   9061   9240
      exon   11452  11597
  Allele     EM:S77127     452    1183   1183   A>G
             source        isSNP  SNP00003080
             source        wetSNP        EM:S77127.v1183.C>T
             consequence   SOD2_link_cds.1     453    Missense     16-16  A>V
  Allele     EM:S77127     452    1456   1456   G>T
             source        wetSNP        EM:S77127.v1456.A>C
             consequence   SOD2_link_cds.1     453    Intron
  Allele     EM:S77127     452    1734   1734   A>G
             source        isSNP  SNP00107369
             consequence   SOD2_link_cds.1     453    Intron
  GIF  SOD2-genomic-fwd.gif


SOD3
Full name : superoxide dismutase 3, extracellular
Link : SOD3_link_cdna
  Subsequence        GB:SOD3      1      1984   #455
  CDS  GB:SOD3.1     723 bp              #456
      ORF    664    1386
  Allele     GB:SOD3       455    835    835    A>G
             source        isSNP  SNP00033027
             consequence   GB:SOD3.1     456    Missense     58-58  T>A
  Allele     GB:SOD3       455    874    874    A>G
             source        isSNP  SNP00062433
             consequence   GB:SOD3.1     456    Silent       71-71  L
  Allele     GB:SOD3       455    1469   1469   A>G
             source        isSNP  SNP00067750
             consequence   GB:SOD3.1     456    3'
  Allele     GB:SOD3       455    1496   1496   A>G
             source        isSNP  SNP00007500
             consequence   GB:SOD3.1     456    3'
  Allele     GB:SOD3       455    1817   1817   G>T
             source        isSNP  SNP00104042
             consequence   GB:SOD3.1     456    3'
  Allele     GB:SOD3       455    1826   1826   A>G
             source        isSNP  SNP00031110
```

TABLE 1 (Cont.)

```
            consequence  GB:SOD3.1    456    3'
  Allele    GB:SOD3      455    1932   1932   A>G
            source       isSNP  SNP00050239
            consequence  GB:SOD3.1    456    3'
  GIF  SOD3-cdna-fwd.gif
Link : FL_1534327_link_genomic
  Subsequence      GB:HSU10116  1      10079  #457
  Subsequence      GB:HSU10116_1534327CD1    5085   5807   #458
  Subsequence      FL_1534327_mrna_build.1   1130   6405   #459
  mRNA      FL_1534327_mrna_build.1  1427 bp     2 exons      #459
       exon  1130   1219
       exon  5069   6405
  CDS  GB:HSU10116_1534327CD1   723 bp       1 exon       #458
       exon  5085   5807
  Allele    GB:HSU10116  457    5256   5256   A>G
            source       isSNP  SNP00033027
            consequence  GB:HSU10116_1534327CD1   458   Missense    58-58  T>A
  Allele    GB:HSU10116  457    5295   5295   A>G
            source       isSNP  SNP00062433
            consequence  GB:HSU10116_1534327CD1   458   Silent      71-71  L
  Allele    GB:HSU10116  457    5890   5890   A>G
            source       isSNP  SNP00067750
            consequence  GB:HSU10116_1534327CD1   458   3'
  Allele    GB:HSU10116  457    5917   5917   A>G
            source       isSNP  SNP00007500
            consequence  GB:HSU10116_1534327CD1   458   3'
  Allele    GB:HSU10116  457    6238   6238   G>T
            source       isSNP  SNP00104042
            consequence  GB:HSU10116_1534327CD1   458   3'
  Allele    GB:HSU10116  457    6247   6247   A>G
            source       isSNP  SNP00031110
            consequence  GB:HSU10116_1534327CD1   458   3'
  Allele    GB:HSU10116  457    6353   6353   A>G
            source       isSNP  SNP00050239
            consequence  GB:HSU10116_1534327CD1   458   3'
  GIF  SOD3-genomic-fwd.gif


SOX9
Full name : SOX9
Link : SOX9_link_cdna
  Subsequence     GB:HSSOX9MRN        1      3923   #460
  CDS  GB:HSSOX9MRN.2      1530 bp            #461
       ORF   360    1889
  Allele     GB:HSSOX9MRN       460    866    866    A>G
             source      isSNP  SNP00092616
             consequence GB:HSSOX9MRN.2      461    Silent       169-169      H
  Allele     GB:HSSOX9MRN       460    1571   1571   A>G
             source      isSNP  SNP00108001
             consequence GB:HSSOX9MRN.2      461    Silent       404-404      P
  Allele     GB:HSSOX9MRN       460    1912   1912   G>T
             source      isSNP  SNP00055269
             consequence GB:HSSOX9MRN.2      461    3'
  Allele     GB:HSSOX9MRN       460    2374   2374   A>G
```

TABLE 1 (Cont.)

```
            source        isSNP  SNP00041454
            consequence  GB:HSSOX9MRN.2       461    3'
  Allele    GB:HSSOX9MRN         460    3224   3224   C>G
            source        isSNP  SNP00061027
            consequence  GB:HSSOX9MRN.2       461    3'
  Allele    GB:HSSOX9MRN         460    3470   3470   A>G
            source        isSNP  SNP00055270
            consequence  GB:HSSOX9MRN.2       461    3'
  GIF  SOX9-cdna-fwd.gif
Link : FL_5425567_link_genomic
  Subsequence        GB:AC007461_8_5425567CD1        63884  60889  #462
  Subsequence        GB:AC007461_8        1       180385        #463
  Subsequence        SOX9_mrna_build.1   64243  58856  #464
  CDS  GB:AC007461_8_5425567CD1          1530 bp       3 exons        #462
       exon    63884  63454
       exon    62557  62304
       exon    61733  60889
  mRNA         SOX9_mrna_build.1   3922 bp       3 exons       #464
       exon    64243  63454
       exon    62557  62304
       exon    61733  58856
  Allele      GB:AC007461_8        463    59309  59309  A>G
              source        isSNP  SNP00055270
              consequence  GB:AC007461_8_5425567CD1        462    3'
  Allele      GB:AC007461_8        463    59555  59555  C>G
              source        isSNP  SNP00061027
              consequence  GB:AC007461_8_5425567CD1        462    3'
  Allele      GB:AC007461_8        463    60078  60078  A>G
              source        isSNP  SNP00010889
              consequence  GB:AC007461_8_5425567CD1        462    3'
  Allele      GB:AC007461_8        463    60404  60404  A>G
              source        isSNP  SNP00041454
              consequence  GB:AC007461_8_5425567CD1        462    3'
  Allele      GB:AC007461_8        463    60866  60866  G>T
              source        isSNP  SNP00055269
              consequence  GB:AC007461_8_5425567CD1        462    3'
  Allele      GB:AC007461_8        463    61207  61207  A>G
              source        isSNP  SNP00108001
              consequence  GB:AC007461_8_5425567CD1        462    Silent
404-404       P
  Allele      GB:AC007461_8        463    62482  62482  A>G
              source        isSNP  SNP00092616
              source        wetSNP        GB:AC007461_8.v62482.G>A
              consequence  GB:AC007461_8_5425567CD1        462    Silent
169-169       H
  GIF  SOX9-genomic-rev.gif


STATI2
Full name : STAT-induced STAT inhibitor-2
Link : FL_2787140_link_cdna
  Subsequence        FN:2787140CB1        1       2587   #465
  CDS  FN:2787140CB1.1     927 bp               #466
       ORF    98     1024
```

TABLE 1 (Cont.)

```
  Allele     FN:2787140CB1       465    1325   1325   A>G
             source        isSNP  SNP00041483
             consequence  FN:2787140CB1.1    466    3'
  Allele     FN:2787140CB1       465    1442   1442   G>T
             source        isSNP  SNP00106962
             consequence  FN:2787140CB1.1    466    3'
  Allele     FN:2787140CB1       465    1470   1470   A>G
             source        isSNP  SNP00041484
             consequence  FN:2787140CB1.1    466    3'
  Allele     FN:2787140CB1       465    1974   1974   A>G
             source        isSNP  SNP00106963
             consequence  FN:2787140CB1.1    466    3'
  GIF STATI2-cdna-fwd.gif
Link : FL_1405668_link_genomic
  Subsequence      GB:AC012085_1       1      177866       #467
  Subsequence      FL_2787140_mrna_build.1  42013  47745  #468
  mRNA       FL_2787140_mrna_build.1  2580 bp      3 exons       #468
      exon   42013  42225
      exon   43694  44045
      exon   45731  47745
  Allele     GB:AC012085_1       467    44268  44268  A>G
             source        isSNP  SNP00070304
  Allele     GB:AC012085_1       467    46492  46492  A>G
             source        isSNP  SNP00041483
  Allele     GB:AC012085_1       467    46609  46609  G>T
             source        isSNP  SNP00106962
  Allele     GB:AC012085_1       467    46637  46637  A>G
             source        isSNP  SNP00041484
  Allele     GB:AC012085_1       467    47141  47141  A>G
             source        isSNP  SNP00106963
  GIF STATI2-genomic-fwd.gif


THBS1
Full name : thrombospondin 1
Link : THBS1_link_cdna
  Subsequence      GB:HSTS       1      5722   #469
  CDS  GB:HSTS.1     3513 bp             #470
      ORF    112    3624
  Allele     GB:HSTS       469    1239   1239   A>G
             source        isSNP  SNP00046537
             consequence  GB:HSTS.1     470    Silent        376-376       D
  Allele     GB:HSTS       469    2210   2210   A>G
             source        isSNP  SNP00046539
             consequence  GB:HSTS.1     470    Missense      700-700       N>S
  Allele     GB:HSTS       469    2979   2979   A>G
             source        isSNP  SNP00061983
             consequence  GB:HSTS.1     470    Silent        956-956       D
  Allele     GB:HSTS       469    3680   3680   G>T
             source        isSNP  SNP00108514
             consequence  GB:HSTS.1     470    3'
  Allele     GB:HSTS       469    3703   3703   A>G
             source        isSNP  SNP00013197
             consequence  GB:HSTS.1     470    3'
```

TABLE 1 (Cont.)

```
  Allele      GB:HSTS       469    3905   3905   A>G
              source        isSNP  SNP00093327
              consequence   GB:HSTS.1     470    3'
  Allele      GB:HSTS       469    5259   5259   A>G
              source        isSNP  SNP00105437
              consequence   GB:HSTS.1     470    3'
  GIF THBS1-cdna-fwd.gif


TIMP1
Full name : Tissue Inhibitor of Metalloproteinase 1
Link : TIMP1_link_cdna
  Subsequence      FN:411388CB1        1      853    #471
  CDS FN:411388CB1.1      621 bp             #472
      ORF    122    742
  Allele      FN:411388CB1       471    365    365    C>G
              source        isSNP  SNP00115174
              consequence   FN:411388CB1.1     472    Missense     82-82  R>G
  GIF TIMP1-cdna-fwd.gif
Link : FL_3013907_link_genomic
  Subsequence      GB:HS230G1   1      125515       #473
  Subsequence      GB:HS230G1_411388CD1      20559  17287  #474
  Subsequence      TIMP1_mrna_build.1        21613  17186  #475
  mRNA        TIMP1_mrna_build.1       843 bp       6 exons      #475
      exon    21613  21501
      exon    20567  20439
      exon    19039  18960
      exon    18770  18644
      exon    18432  18308
      exon    17454  17186
  CDS GB:HS230G1_411388CD1      621 bp       5 exons      #474
      exon    20559  20439
      exon    19039  18960
      exon    18770  18644
      exon    18432  18308
      exon    17454  17287
  Allele      GB:HS230G1    473    17434  17434  A>G
              source        wetSNP        GB:HS230G1.v17434.G>A
              consequence   GB:HS230G1_411388CD1      474    Silent       158-158
      I
  Allele      GB:HS230G1    473    17550  17550  A>G
              source        isSNP  SNP00099224
              consequence   GB:HS230G1_411388CD1      474    Intron
  Allele      GB:HS230G1    473    18046  18046  A>G
              source        isSNP  SNP00099223
              consequence   GB:HS230G1_411388CD1      474    Intron
  Allele      GB:HS230G1    473    18088  18088  A>G
              source        isSNP  SNP00030937
              consequence   GB:HS230G1_411388CD1      474    Intron
  Allele      GB:HS230G1    473    18389  18389  A>G
              source        wetSNP        GB:HS230G1.v18389.A>G
              consequence   GB:HS230G1_411388CD1      474    Silent       124-124
      F
  Allele      GB:HS230G1    473    18495  18495  C>G
```

TABLE 1 (Cont.)

```
            source        isSNP  SNP00099222
            source        wetSNP        GB:HS230G1.v18495.C>G
            consequence   GB:HS230G1_411388CD1      474    Intron
  Allele    GB:HS230G1    473    18711  18711  A>G
            source        wetSNP        GB:HS230G1.v18711.G>A
            consequence   GB:HS230G1_411388CD1      474    Silent       87-87  P
  Allele    GB:HS230G1    473    18728  18728  C>G
            source        isSNP  SNP00115174
            consequence   GB:HS230G1_411388CD1      474    Missense     82-82  R>G
  GIF  TIMP1-genomic-rev.gif


TIMP2
Full name : Tissue Inhibitor of Metalloproteinase-2.
Link : TIMP2_link_genomic
  Subsequence       TIMP2_cds.1  822    3126    #476
  Subsequence       GB:S68860_1  1      970     #477
  Subsequence       GB:U44382_1  1071   1320    #478
  Subsequence       GB:U44383_1  1421   1644    #479
  Subsequence       GB:U44384_1  1745   2283    #480
  Subsequence       GB:U44385_1  2384   3750    #481
  Subsequence       TIMP2_mrna_build.1          810    3251   #482
  CDS  TIMP2_cds.1  663 bp       5 exons        #476
       exon   822    951
       exon   1125   1225
       exon   1504   1612
       exon   1939   2063
       exon   2929   3126
  mRNA        TIMP2_mrna_build.1        800 bp       5 exons       #482
       exon   810    951
       exon   1125   1225
       exon   1504   1612
       exon   1939   2063
       exon   2929   3251
  Allele      GB:U44383_1  479    155    155    A>G
              source       wetSNP        GB:U44383_1.v155.G>A
              consequence  TIMP2_cds.1  476    Silent       101-101      S
  GIF  TIMP2-genomic-fwd.gif


TNA
Full name : tetranectin
Link : TNA_link_cdna
  Subsequence       GB:NM_003278        1      874    #483
  CDS  GB:NM_003278.1      609 bp              #484
       ORF    94     702
  Allele      GB:NM_003278         483    409    409    A>G
              source       isSNP  SNP00007942
              consequence  GB:NM_003278.1      484    Missense     106-106      S>G
  Allele      GB:NM_003278         483    744    744    A>G
              source       isSNP  SNP00007943
              consequence  GB:NM_003278.1      484    3'
  GIF  TNA-cdna-fwd.gif
```

TABLE 1 (Cont.)

```
Link : TNA_link_genomic
  Subsequence       TNA_cds.1    254    1629   #485
  Subsequence       TNA_cds.2    254    1629   #486
  Subsequence       GB:X70910_1  1      570    #487
  Subsequence       GB:X70911_1  671    978    #488
  Subsequence       GB:X70912_1  1079   1805   #489
  Subsequence       TNA_mrna_build.1    164    1776   #490
  CDS TNA_cds.1     609 bp       3 exons       #485
      exon   254    362
      exon   829    927
      exon   1229   1629
  CDS TNA_cds.2     510 bp       2 exons       #486
      exon   254    362
      exon   1229   1629
  mRNA        TNA_mrna_build.1   846 bp        3 exons       #490
      exon   164    362
      exon   829    927
      exon   1229   1776
  Allele      GB:X70912_1  489    258    258    A>G
              source       isSNP  SNP00007942
              consequence  TNA_cds.1    485    Missense     106-106      S>G
              consequence  TNA_cds.2    486    Missense     73-73  S>G
  Allele      GB:X70912_1  489    593    593    A>G
              source       isSNP  SNP00007943
              consequence  TNA_cds.1    485    3'
              consequence  TNA_cds.2    486    3'
  GIF TNA-genomic-fwd.gif


TNFAIP6
Full name : tumor necrosis factor, alpha-induced protein 6
Link : TNFAIP6_link_cdna
  Subsequence       GB:NM_007115_1      1      1414   #491
  CDS GB:NM_007115_1.1   834 bp             #492
      ORF    69     902
  Allele      GB:NM_007115_1      491    499    499    A>G
              source       isSNP  SNP00040822
              consequence  GB:NM_007115_1.1   492    Missense     144-144      R>Q
  Allele      GB:NM_007115_1      491    1143   1143   C>G
              source       isSNP  SNP00040823
              consequence  GB:NM_007115_1.1   492    3'
  GIF TNFAIP6-cdna-fwd.gif
Link : FL_1000909_link_genomic
  Subsequence       GB:AC009311_1_191918CD1   132384      154250      #493
  Subsequence       GB:AC009311_1       1      160198      #494
  Subsequence       TNFAIP6_mrna_build.1      132314      154760      #495
  mRNA        TNFAIP6_mrna_build.1      1414 bp      6 exons      #495
      exon   132314       132477
      exon   138660       138797
      exon   140773       140934
      exon   144737       144965
      exon   148266       148306
      exon   154081       154760
  CDS GB:AC009311_1_191918CD1  834 bp       6 exons      #493
```

TABLE 1 (Cont.)

```
        exon   132384        132477
        exon   138660        138797
        exon   140773        140934
        exon   144737        144965
        exon   148266        148306
        exon   154081        154250
   Allele      GB:AC009311_1        494    140934        140934        A>G
               source        wetSNP        GB:AC009311_1.v140934.G>A
               consequence   GB:AC009311_1_191918CD1   493    Missense     132-132
        A>T
   Allele      GB:AC009311_1        494    140942        140942        A>T
               source        wetSNP        GB:AC009311_1.v140942.A>T
               consequence   GB:AC009311_1_191918CD1   493    Intron
   Allele      GB:AC009311_1        494    144773        144773        A>G
               source        isSNP  SNP00040822
               source        wetSNP        GB:AC009311_1.v144773.A>G
               consequence   GB:AC009311_1_191918CD1   493    Missense     144-144
        Q>R
   Allele      GB:AC009311_1        494    148030        148030        A>G
               source        dbSNP  gnl|dbSNP|ss645109_allele
               consequence   GB:AC009311_1_191918CD1   493    Intron
   Allele      GB:AC009311_1        494    148229        148229        A>G
               source        wetSNP        GB:AC009311_1.v148229.T>C
               consequence   GB:AC009311_1_191918CD1   493    Intron
   Allele      GB:AC009311_1        494    148245        148245        A>G
               source        wetSNP        GB:AC009311_1.v148245.T>C
               consequence   GB:AC009311_1_191918CD1   493    Intron
   Allele      GB:AC009311_1        494    154493        154493        C>G
               source        isSNP  SNP00040823
               consequence   GB:AC009311_1_191918CD1   493    3'
   GIF  TNFAIP6-genomic-fwd.gif


TNFRSF11B
Full name : TNFRSF11B
Link : TNFRSF11B_link_cdna
  Subsequence       GB:AB002146  1       1206    #496
  CDS  GB:AB002146.1        1206  bp             #497
       ORF    1      1206
  Allele      GB:AB002146  496    768    768    A>G
              source        isSNP  SNP00028816
              consequence  GB:AB002146.1        497    Silent        256-256      L
  GIF  TNFRSF11B-cdna-fwd.gif
Link : TNFRSF11B_link_genomic
  Subsequence       TNFRSF11B_cds.1     125     9057    #498
  Subsequence       GB:E15270_1  1      9898    #499
  CDS  TNFRSF11B_cds.1      1176  bp       4  exons       #498
       exon   130    499
       exon   4504   4695
       exon   6716   6940
       exon   8669   9057
  Allele      GB:E15270_1  499    503     503    A>G
              source        wetSNP        GB:E15270_1.v503.C>T
              consequence  TNFRSF11B_cds.1      498    Intron
```

TABLE 1 (Cont.)

Allele	GB:E15270_1	499	4499	4499	A>G		
	source	wetSNP		GB:E15270_1.v4499.C>T			
	consequence	TNFRSF11B_cds.1		498	Intron		
Allele	GB:E15270_1	499	4661	4661	A>G		
	source	wetSNP		GB:E15270_1.v4661.C>T			
	consequence	TNFRSF11B_cds.1		498	Silent	176-176	S
Allele	GB:E15270_1	499	4749	4752	TCTG>TG		
	source	wetSNP		GB:E15270_1.v4749.TCTG>TG			
	consequence	TNFRSF11B_cds.1		498	Intron		
Allele	GB:E15270_1	499	6599	6599	A>G		
	source	wetSNP		GB:E15270_1.v6599.G>A			
	consequence	TNFRSF11B_cds.1		498	Intron		
Allele	GB:E15270_1	499	6837	6837	A>G		
	source	wetSNP		GB:E15270_1.v6837.G>A			
	consequence	TNFRSF11B_cds.1		498	Silent	228-228	E
Allele	GB:E15270_1	499	6891	6891	A>G		
	source	isSNP SNP00028816					
	consequence	TNFRSF11B_cds.1		498	Silent	246-246	L

GIF TNFRSF11B-genomic-fwd.gif

TABLE 2

Gene	Sequence	Seq Offset	Pol Position	DNA change	Qualifier	Peptide change	Codon	No. ind screened	Freq of variant	No. AA	No. AB	No. BB	Ethnic group
AACT	GB:AL049839_2	59566	148	G > A	MIS	A>T	9	47	0.5	14	19	14	all
AACT	GB:AL049839_2	59566	148	G > A	MIS	A>T	9	10	0.2	7	2	1	AFA
AACT	GB:AL049839_2	59566	148	G > A	MIS	A>T	9	13	0.7308	0	7	6	ASI
AACT	GB:AL049839_2	59566	148	G > A	MIS	A>T	9	17	0.6471	1	10	6	CAU
AACT	GB:AL049839_2	59566	148	G > A	MIS	A>T	9	7	0.1429	6	0	1	HIS
AACT	GB:AL049839_2	68882	55	A > G	MIS	M>V	414	47	0.0106	46	1	0	all
AACT	GB:AL049839_2	68882	55	A > G	MIS	M>V	414	10	0	10	0	0	AFA
AACT	GB:AL049839_2	68882	55	A > G	MIS	M>V	414	13	0.0385	12	1	0	ASI
AACT	GB:AL049839_2	68882	55	A > G	MIS	M>V	414	17	0	17	0	0	CAU
AACT	GB:AL049839_2	68882	55	A > G	MIS	M>V	414	7	0	7	0	0	HIS
ABL1	GB:U07563_1	58876	85	C > G	NCD			47	0.0106	46	1	0	all
ABL1	GB:U07563_1	58876	85	C > G	NCD			10	0	10	0	0	AFA
ABL1	GB:U07563_1	58876	85	C > G	NCD			13	0.0385	12	1	0	ASI
ABL1	GB:U07563_1	58876	85	C > G	NCD			17	0	17	0	0	CAU
ABL1	GB:U07563_1	58876	85	C > G	NCD			7	0	7	0	0	HIS
ABL1	GB:U07563_1	68640	121	T > C	NCD			47	0.0319	44	3	0	all
ABL1	GB:U07563_1	68640	121	T > C	NCD			10	0.15	7	3	0	AFA
ABL1	GB:U07563_1	68640	121	T > C	NCD			13	0	13	0	0	ASI
ABL1	GB:U07563_1	68640	121	T > C	NCD			17	0	17	0	0	CAU
ABL1	GB:U07563_1	68640	121	T > C	NCD			7	0	7	0	0	HIS
ABL1	GB:U07563_1	74901	143	A > G	SIL	E	499	47	0.0745	40	7	0	all
ABL1	GB:U07563_1	74901	143	A > G	SIL	E	499	10	0.1	8	2	0	AFA
ABL1	GB:U07563_1	74901	143	A > G	SIL	E	499	13	0	13	0	0	ASI
ABL1	GB:U07563_1	74901	143	A > G	SIL	E	499	17	0.1176	13	4	0	CAU
ABL1	GB:U07563_1	74901	143	A > G	SIL	E	499	7	0.0714	6	1	0	HIS
ABL1	GB:U07563_1	78921	87	G > A	SIL	E	623	46	0.0435	42	4	0	all
ABL1	GB:U07563_1	78921	87	G > A	SIL	E	623	10	0.2	6	4	0	AFA
ABL1	GB:U07563_1	78921	87	G > A	SIL	E	623	13	0	13	0	0	ASI
ABL1	GB:U07563_1	78921	87	G > A	SIL	E	623	17	0	17	0	0	CAU
ABL1	GB:U07563_1	78921	87	G > A	SIL	E	623	6	0	6	0	0	HIS
ABL1	GB:U07563_1	79239	136	G > A	SIL	T	729	47	0.0106	46	1	0	all
ABL1	GB:U07563_1	79239	136	G > A	SIL	T	729	10	0.05	9	1	0	AFA
ABL1	GB:U07563_1	79239	136	G > A	SIL	T	729	13	0	13	0	0	ASI
ABL1	GB:U07563_1	79239	136	G > A	SIL	T	729	17	0	17	0	0	CAU
ABL1	GB:U07563_1	79239	136	G > A	SIL	T	729	7	0	7	0	0	HIS
ABL1	GB:U07563_1	79404	125	C > G	SIL	P	784	46	0.0761	39	7	0	all
ABL1	GB:U07563_1	79404	125	C > G	SIL	P	784	10	0.15	7	3	0	AFA
ABL1	GB:U07563_1	79404	125	C > G	SIL	P	784	13	0	13	0	0	ASI
ABL1	GB:U07563_1	79404	125	C > G	SIL	P	784	16	0.0938	13	3	0	CAU
ABL1	GB:U07563_1	79404	125	C > G	SIL	P	784	7	0.0714	6	1	0	HIS

TABLE 2

Gene	Sequence	Seq Offset	Pol Position	DNA change	Qualifier	Peptide change	Codon	No. ind screened	Freq of variant	No. AA	No. AB	No. BB	Ethnic group
ABL1	GB:U07563_1	79657	171	C > T	MIS	P>S	869	47	0.0106	46	1	0	all
ABL1	GB:U07563_1	79657	171	C > T	MIS	P>S	869	10	0	10	0	0	AFA
ABL1	GB:U07563_1	79657	171	C > T	MIS	P>S	869	13	0	13	0	0	ASI
ABL1	GB:U07563_1	79657	171	C > T	MIS	P>S	869	17	0.0294	16	1	0	CAU
ABL1	GB:U07563_1	79657	171	C > T	MIS	P>S	869	7	0	7	0	0	HIS
ABL1	GB:U07563_1	79750	76	C > T	MIS	P>S	900	46	0.0109	45	1	0	all
ABL1	GB:U07563_1	79750	76	C > T	MIS	P>S	900	10	0.05	9	1	0	AFA
ABL1	GB:U07563_1	79750	76	C > T	MIS	P>S	900	13	0	13	0	0	ASI
ABL1	GB:U07563_1	79750	76	C > T	MIS	P>S	900	16	0	16	0	0	CAU
ABL1	GB:U07563_1	79750	76	C > T	MIS	P>S	900	7	0	7	0	0	HIS
ABL1	GB:U07563_1	80376	70	G > A	SIL	P	1108	47	0.4894	1	46	0	all
ABL1	GB:U07563_1	80376	70	G > A	SIL	P	1108	10	0.5	0	10	0	AFA
ABL1	GB:U07563_1	80376	70	G > A	SIL	P	1108	13	0.5	0	13	0	ASI
ABL1	GB:U07563_1	80376	70	G > A	SIL	P	1108	17	0.4706	1	16	0	CAU
ABL1	GB:U07563_1	80376	70	G > A	SIL	P	1108	7	0.5	0	7	0	HIS
ACLP	GB:AC006454_3	141107	197	GA > GGA	NCD			44	0.0114	43	1	0	all
ACLP	GB:AC006454_3	141107	197	GA > GGA	NCD			9	0	9	0	0	AFA
ACLP	GB:AC006454_3	141107	197	GA > GGA	NCD			12	0	12	0	0	ASI
ACLP	GB:AC006454_3	141107	197	GA > GGA	NCD			17	0.0294	16	1	0	CAU
ACLP	GB:AC006454_3	141107	197	GA > GGA	NCD			6	0	6	0	0	HIS
ACLP	GB:AC006454_3	142383	132	T > C	MIS	T>A	424	47	0.0106	46	1	0	all
ACLP	GB:AC006454_3	142383	132	T > C	MIS	T>A	424	10	0	10	0	0	AFA
ACLP	GB:AC006454_3	142383	132	T > C	MIS	T>A	424	13	0	13	0	0	ASI
ACLP	GB:AC006454_3	142383	132	T > C	MIS	T>A	424	17	0.0294	16	1	0	CAU
ACLP	GB:AC006454_3	142383	132	T > C	MIS	T>A	424	7	0	7	0	0	HIS
ACLP	GB:AC006454_3	145005	21	C > T	NCD			46	0.0217	44	2	0	all
ACLP	GB:AC006454_3	145005	21	C > T	NCD			10	0	10	0	0	AFA
ACLP	GB:AC006454_3	145005	21	C > T	NCD			12	0	12	0	0	ASI
ACLP	GB:AC006454_3	145005	21	C > T	NCD			17	0.0588	15	2	0	CAU
ACLP	GB:AC006454_3	145005	21	C > T	NCD			7	0	7	0	0	HIS
ANK	GB:AC026437_2	25779	182	C > T	SIL	A	51	47	0.117	37	9	1	all
ANK	GB:AC026437_2	25779	182	C > T	SIL	A	51	10	0.15	7	3	0	AFA
ANK	GB:AC026437_2	25779	182	C > T	SIL	A	51	13	0	13	0	0	ASI
ANK	GB:AC026437_2	25779	182	C > T	SIL	A	51	17	0.2059	11	5	1	CAU
ANK	GB:AC026437_2	25779	182	C > T	SIL	A	51	7	0.0714	6	1	0	HIS
ANK	GB:AC026437_2	25807	210	G > A	NCD			47	0.0745	41	5	1	all
ANK	GB:AC026437_2	25807	210	G > A	NCD			10	0	10	0	0	AFA
ANK	GB:AC026437_2	25807	210	G > A	NCD			13	0	13	0	0	ASI
ANK	GB:AC026437_2	25807	210	G > A	NCD			17	0.1765	12	4	1	CAU
ANK	GB:AC026437_2	25807	210	G > A	NCD			7	0.0714	6	1	0	HIS

TABLE 2

Gene	Sequence	Seq Offset	Pol Position	DNA change	Qualifier	Peptide change	Codon	No. ind screened	Freq of variant	No. AA	No. AB	No. BB	Ethnic group
ANK	GB:AC026437_2	36172	24	T > C	NCD			46	0.0109	45	1	0	all
ANK	GB:AC026437_2	36172	24	T > C	NCD			10	0	10	0	0	AFA
ANK	GB:AC026437_2	36172	24	T > C	NCD			13	0.0385	12	1	0	ASI
ANK	GB:AC026437_2	36172	24	T > C	NCD			17	0	17	0	0	CAU
ANK	GB:AC026437_2	36172	24	T > C	NCD			6	0	6	0	0	HIS
ANK	GB:AC026437_2	52817	21	C > A	NCD			47	0.0106	46	1	0	all
ANK	GB:AC026437_2	52817	21	C > A	NCD			10	0	10	0	0	AFA
ANK	GB:AC026437_2	52817	21	C > A	NCD			13	0.0385	12	1	0	ASI
ANK	GB:AC026437_2	52817	21	C > A	NCD			17	0	17	0	0	CAU
ANK	GB:AC026437_2	52817	21	C > A	NCD			7	0	7	0	0	HIS
ANK	GB:AC026437_2	52899	103	A > G	SIL	A	274	47	0.0851	40	6	1	all
ANK	GB:AC026437_2	52899	103	A > G	SIL	A	274	10	0.05	9	1	0	AFA
ANK	GB:AC026437_2	52899	103	A > G	SIL	A	274	13	0.1538	10	2	1	ASI
ANK	GB:AC026437_2	52899	103	A > G	SIL	A	274	17	0.0294	16	1	0	CAU
ANK	GB:AC026437_2	52899	103	A > G	SIL	A	274	7	0.1429	5	2	0	HIS
ANK	GB:AC026437_2	52899	40	A > G	SIL	A	274	47	0.0851	40	6	1	all
ANK	GB:AC026437_2	52899	40	A > G	SIL	A	274	10	0.05	9	1	0	AFA
ANK	GB:AC026437_2	52899	40	A > G	SIL	A	274	13	0.1538	10	2	1	ASI
ANK	GB:AC026437_2	52899	40	A > G	SIL	A	274	17	0.0294	16	1	0	CAU
ANK	GB:AC026437_2	52899	40	A > G	SIL	A	274	7	0.1429	5	2	0	HIS
ANK	GB:AC026437_2	52962	103	T > G	NCD			47	0.0745	40	7	0	all
ANK	GB:AC026437_2	52962	103	T > G	NCD			10	0.05	9	1	0	AFA
ANK	GB:AC026437_2	52962	103	T > G	NCD			13	0	13	0	0	ASI
ANK	GB:AC026437_2	52962	103	T > G	NCD			17	0.1176	13	4	0	CAU
ANK	GB:AC026437_2	52962	103	T > G	NCD			7	0.1429	5	2	0	HIS
ANK	GB:AC026437_2	81235	144	T > C	NCD			47	0.0106	46	1	0	all
ANK	GB:AC026437_2	81235	144	T > C	NCD			10	0.05	9	1	0	AFA
ANK	GB:AC026437_2	81235	144	T > C	NCD			13	0	13	0	0	ASI
ANK	GB:AC026437_2	81235	144	T > C	NCD			17	0	17	0	0	CAU
ANK	GB:AC026437_2	81235	144	T > C	NCD			7	0	7	0	0	HIS
ANK	GB:AC026437_2	83587	186	G > A	NCD			47	0.0106	46	1	0	all
ANK	GB:AC026437_2	83587	186	G > A	NCD			10	0	10	0	0	AFA
ANK	GB:AC026437_2	83587	186	G > A	NCD			13	0.0385	12	1	0	ASI
ANK	GB:AC026437_2	83587	186	G > A	NCD			17	0	17	0	0	CAU
ANK	GB:AC026437_2	83587	186	G > A	NCD			7	0	7	0	0	HIS
ANK	GB:AC026437_2	83607	206	A > G	NCD			47	0.0745	40	7	0	all
ANK	GB:AC026437_2	83607	206	A > G	NCD			10	0.2	6	4	0	AFA
ANK	GB:AC026437_2	83607	206	A > G	NCD			13	0	13	0	0	ASI
ANK	GB:AC026437_2	83607	206	A > G	NCD			17	0.0588	15	2	0	CAU
ANK	GB:AC026437_2	83607	206	A > G	NCD			7	0.0714	6	1	0	HIS

TABLE 2

Gene	Sequence	Seq Offset	Pol Position	DNA change	Qualifier	Peptide change	Codon	No. ind screened	Freq of variant	No. AA	No. AB	No. BB	Ethnic group
BGLAP	GB:AC007227_2	34618	75	G > C	SIL	A	92	47	0.0319	44	3	0	all
BGLAP	GB:AC007227_2	34618	75	G > C	SIL	A	92	10	0.15	7	3	0	AFA
BGLAP	GB:AC007227_2	34618	75	G > C	SIL	A	92	13	0	13	0	0	ASI
BGLAP	GB:AC007227_2	34618	75	G > C	SIL	A	92	17	0	17	0	0	CAU
BGLAP	GB:AC007227_2	34618	75	G > C	SIL	A	92	7	0	7	0	0	HIS
BGLAP	GB:AC007227_2	34977	79	G > T	MIS	Q>K	40	47	0.0319	44	3	0	all
BGLAP	GB:AC007227_2	34977	79	G > T	MIS	Q>K	40	10	0	10	0	0	AFA
BGLAP	GB:AC007227_2	34977	79	G > T	MIS	Q>K	40	13	0.1154	10	3	0	ASI
BGLAP	GB:AC007227_2	34977	79	G > T	MIS	Q>K	40	17	0	17	0	0	CAU
BGLAP	GB:AC007227_2	34977	79	G > T	MIS	Q>K	40	7	0	7	0	0	HIS
BGN	GB:U82695	18161	56	A > G	SIL	E	40	47	0.0106	46	1	0	all
BGN	GB:U82695	18161	56	A > G	SIL	E	40	10	0.05	9	1	0	AFA
BGN	GB:U82695	18161	56	A > G	SIL	E	40	13	0	13	0	0	ASI
BGN	GB:U82695	18161	56	A > G	SIL	E	40	17	0	17	0	0	CAU
BGN	GB:U82695	18161	56	A > G	SIL	E	40	7	0	7	0	0	HIS
BGN	GB:U82695	18182	77	G > A	SIL	S	47	47	0.4681	17	16	14	all
BGN	GB:U82695	18182	77	G > A	SIL	S	47	10	0.4	4	4	2	AFA
BGN	GB:U82695	18182	77	G > A	SIL	S	47	13	0.5385	5	2	6	ASI
BGN	GB:U82695	18182	77	G > A	SIL	S	47	17	0.4706	5	8	4	CAU
BGN	GB:U82695	18182	77	G > A	SIL	S	47	7	0.4286	3	2	2	HIS
BGN	GB:U82695	18330	225	G > A	NCD			47	0.0319	45	1	1	all
BGN	GB:U82695	18330	225	G > A	NCD			10	0.05	9	1	0	AFA
BGN	GB:U82695	18330	225	G > A	NCD			13	0	13	0	0	ASI
BGN	GB:U82695	18330	225	G > A	NCD			17	0.0588	16	0	1	CAU
BGN	GB:U82695	18330	225	G > A	NCD			7	0	7	0	0	HIS
BGN	GB:U82695	18354	249	G > A	NCD			47	0:1064	42	0	5	all
BGN	GB:U82695	18354	249	G > A	NCD			10	0.2	8	0	2	AFA
BGN	GB:U82695	18354	249	G > A	NCD			13	0.0769	12	0	1	ASI
BGN	GB:U82695	18354	249	G > A	NCD			17	0.0588	16	0	1	CAU
BGN	GB:U82695	18354	249	G > A	NCD			7	0.1429	6	0	1	HIS
BGN	GB:U82695	19460	230	T > C	SIL	S	180	46	0.6957	9	10	27	all
BGN	GB:U82695	19460	230	T > C	SIL	S	180	9	0.8333	1	1	7	AFA
BGN	GB:U82695	19460	230	T > C	SIL	S	180	13	0.6923	3	2	8	ASI
BGN	GB:U82695	19460	230	T > C	SIL	S	180	17	0.6765	3	5	9	CAU
BGN	GB:U82695	19460	230	T > C	SIL	S	180	7	0.5714	2	2	3	HIS
BGN	GB:U82695	21566	22	G > T	NCD			47	0.1383	36	9	2	all
BGN	GB:U82695	21566	22	G > T	NCD			10	0.05	9	1	0	AFA
BGN	GB:U82695	21566	22	G > T	NCD			13	0.0769	12	0	1	ASI
BGN	GB:U82695	21566	22	G > T	NCD			17	0.2353	10	6	1	CAU
BGN	GB:U82695	21566	22	G > T	NCD			7	0.1429	5	2	0	HIS

TABLE 2

Gene	Sequence	Seq Offset	Pol Position	DNA change	Qualifier	Peptide change	Codon	No. Ind screened	Freq of variant	No. AA	No. AB	No. BB	Ethnic group
BGN	GB:U82695	21639	95	C > T	NCD			47	0.0426	44	2	1	all
BGN	GB:U82695	21639	95	C > T	NCD			10	0	10	0	0	AFA
BGN	GB:U82695	21639	95	C > T	NCD			13	0.1538	10	2	1	ASI
BGN	GB:U82695	21639	95	C > T	NCD			17	0	17	0	0	CAU
BGN	GB:U82695	21639	95	C > T	NCD			7	0	7	0	0	HIS
BMP2	GB:HS859D4	168341	121	T > A	MIS	R>S	190	47	0.766	4	14	29	all
BMP2	GB:HS859D4	168341	121	T > A	MIS	R>S	190	10	0.8	0	4	6	AFA
BMP2	GB:HS859D4	168341	121	T > A	MIS	R>S	190	13	0.7692	2	2	9	ASI
BMP2	GB:HS859D4	168341	121	T > A	MIS	R>S	190	17	0.7059	2	6	9	CAU
BMP2	GB:HS859D4	168341	121	T > A	MIS	R>S	190	7	0.8571	0	2	5	HIS
BMP4	GB:HSU43842	9215	189	C > T	MIS	A>V	152	43	0.3721	27	0	16	all
BMP4	GB:HSU43842	9215	189	C > T	MIS	A>V	152	8	0.5	4	0	4	AFA
BMP4	GB:HSU43842	9215	189	C > T	MIS	A>V	152	11	0.5455	5	0	6	ASI
BMP4	GB:HSU43842	9215	189	C > T	MIS	A>V	152	17	0.2353	13	0	4	CAU
BMP4	GB:HSU43842	9215	189	C > T	MIS	A>V	152	7	0.2857	5	0	2	HIS
BMP4	GB:HSU43842	9215	30	C > T	MIS	A>V	152	47	0.766	11	0	36	all
BMP4	GB:HSU43842	9215	30	C > T	MIS	A>V	152	10	0.9	1	0	9	AFA
BMP4	GB:HSU43842	9215	30	C > T	MIS	A>V	152	13	0.8462	2	0	11	ASI
BMP4	GB:HSU43842	9215	30	C > T	MIS	A>V	152	17	0.7059	5	0	12	CAU
BMP4	GB:HSU43842	9215	30	C > T	MIS	A>V	152	7	0.5714	3	0	4	HIS
CBFA1	GB:HSCBFA1S3	177	88	C > T	SIL	N	183	46	0.0109	45	1	0	all
CBFA1	GB:HSCBFA1S3	177	88	C > T	SIL	N	183	10	0.05	9	1	0	AFA
CBFA1	GB:HSCBFA1S3	177	88	C > T	SIL	N	183	13	0	13	0	0	ASI
CBFA1	GB:HSCBFA1S3	177	88	C > T	SIL	N	183	17	0	17	0	0	CAU
CBFA1	GB:HSCBFA1S3	177	88	C > T	SIL	N	183	6	0	6	0	0	HIS
CBFA1	GB:HSCBFA1S8	490	42	C > T	SIL	S	503	46	0.0109	45	1	0	all
CBFA1	GB:HSCBFA1S8	490	42	C > T	SIL	S	503	9	0	9	0	0	AFA
CBFA1	GB:HSCBFA1S8	490	42	C > T	SIL	S	503	13	0.0385	12	1	0	ASI
CBFA1	GB:HSCBFA1S8	490	42	C > T	SIL	S	503	17	0	17	0	0	CAU
CBFA1	GB:HSCBFA1S8	490	42	C > T	SIL	S	503	7	0	7	0	0	HIS
CD36	EM:HSCD36G10	92	74	T > C	SIL	F	293	46	0.0109	45	1	0	all
CD36	EM:HSCD36G10	92	74	T > C	SIL	F	293	10	0	10	0	0	AFA
CD36	EM:HSCD36G10	92	74	T > C	SIL	F	293	13	0.0385	12	1	0	ASI
CD36	EM:HSCD36G10	92	74	T > C	SIL	F	293	16	0	16	0	0	CAU
CD36	EM:HSCD36G10	92	74	T > C	SIL	F	293	7	0	7	0	0	HIS
CD68	GB:AC007421_12	90707	123	C > T	MIS	A>T	340	47	0.0106	46	1	0	all
CD68	GB:AC007421_12	90707	123	C > T	MIS	A>T	340	10	0	10	0	0	AFA
CD68	GB:AC007421_12	90707	123	C > T	MIS	A>T	340	13	0	13	0	0	ASI
CD68	GB:AC007421_12	90707	123	C > T	MIS	A>T	340	17	0.0294	16	1	0	CAU
CD68	GB:AC007421_12	90707	123	C > T	MIS	A>T	340	7	0	7	0	0	HIS

TABLE 2

Gene	Sequence	Seq Offset	Pol Position	DNA change	Qualifier	Peptide change	Codon	No. ind screened	Freq of variant	No. AA	No. AB	No. BB	Ethnic group
CD68	GB:AC007421_12	91388	62	G > T	SPL			47	0.1064	37	10	0	all
CD68	GB:AC007421_12	91388	62	G > T	SPL			10	0.05	9	1	0	AFA
CD68	GB:AC007421_12	91388	62	G > T	SPL			13	0.1538	9	4	0	ASI
CD68	GB:AC007421_12	91388	62	G > T	SPL			17	0.1176	13	4	0	CAU
CD68	GB:AC007421_12	91388	62	G > T	SPL			7	0.0714	6	1	0	HIS
CILP	GB:AB022430_1	3567	39	A > C	NCD			47	0.0106	46	1	0	all
CILP	GB:AB022430_1	3567	39	A > C	NCD			10	0.05	9	1	0	AFA
CILP	GB:AB022430_1	3567	39	A > C	NCD			13	0	13	0	0	ASI
CILP	GB:AB022430_1	3567	39	A > C	NCD			17	0	17	0	0	CAU
CILP	GB:AB022430_1	3567	39	A > C	NCD			7	0	7	0	0	HIS
CILP	GB:AB022430_1	11286	61	T > A	NCD			47	0.0106	46	1	0	all
CILP	GB:AB022430_1	11286	61	T > A	NCD			10	0.05	9	1	0	AFA
CILP	GB:AB022430_1	11286	61	T > A	NCD			13	0	13	0	0	ASI
CILP	GB:AB022430_1	11286	61	T > A	NCD			17	0	17	0	0	CAU
CILP	GB:AB022430_1	11286	61	T > A	NCD			7	0	7	0	0	HIS
CILP	GB:AB022430_1	11491	71	C > T	MIS	T>I	395	47	0.5106	16	14	17	all
CILP	GB:AB022430_1	11491	71	C > T	MIS	T>I	395	10	0.25	5	5	0	AFA
CILP	GB:AB022430_1	11491	71	C > T	MIS	T>I	395	13	0.8462	1	2	10	ASI
CILP	GB:AB022430_1	11491	71	C > T	MIS	T>I	395	17	0.5	6	5	6	CAU
CILP	GB:AB022430_1	11491	71	C > T	MIS	T>I	395	7	0.2857	4	2	1	HIS
CILP	GB:AB022430_1	14421	136	G > C	MIS	R>T	446	47	0.0106	46	1	0	all
CILP	GB:AB022430_1	14421	136	G > C	MIS	R>T	446	10	0	10	0	0	AFA
CILP	GB:AB022430_1	14421	136	G > C	MIS	R>T	446	13	0.0385	12	1	0	ASI
CILP	GB:AB022430_1	14421	136	G > C	MIS	R>T	446	17	0	17	0	0	CAU
CILP	GB:AB022430_1	14421	136	G > C	MIS	R>T	446	7	0	7	0	0	HIS
CILP	GB:AB022430_1	15116	122	G > A	MIS	V>M	678	47	0.0106	46	1	0	all
CILP	GB:AB022430_1	15116	122	G > A	MIS	V>M	678	10	0	10	0	0	AFA
CILP	GB:AB022430_1	15116	122	G > A	MIS	V>M	678	13	0.0385	12	1	0	ASI
CILP	GB:AB022430_1	15116	122	G > A	MIS	V>M	678	17	0	17	0	0	CAU
CILP	GB:AB022430_1	15116	122	G > A	MIS	V>M	678	7	0	7	0	0	HIS
CILP	GB:AB022430_1	15670	146	G > A	SIL	T	862	47	0.0106	46	1	0	all
CILP	GB:AB022430_1	15670	146	G > A	SIL	T	862	10	0.05	9	1	0	AFA
CILP	GB:AB022430_1	15670	146	G > A	SIL	T	862	13	0	13	0	0	ASI
CILP	GB:AB022430_1	15670	146	G > A	SIL	T	862	17	0	17	0	0	CAU
CILP	GB:AB022430_1	15670	146	G > A	SIL	T	862	7	0	7	0	0	HIS
CTSC	GB:AC011088_8	124932	57	A > T	NCD			47	0.0106	46	1	0	all
CTSC	GB:AC011088_8	124932	57	A > T	NCD			10	0.05	9	1	0	AFA
CTSC	GB:AC011088_8	124932	57	A > T	NCD			13	0	13	0	0	ASI
CTSC	GB:AC011088_8	124932	57	A > T	NCD			17	0	17	0	0	CAU
CTSC	GB:AC011088_8	124932	57	A > T	NCD			7	0	7	0	0	HIS

TABLE 2

Gene	Sequence	Seq Offset	Pol Position	DNA change	Qualifier	Peptide change	Codon	No. ind screened	Freq of variant	No. AA	No. AB	No. BB	Ethnic group
CTSC	GB:AC011088_8	125028	153	A > G	MIS	I>T	153	47	0.9149	0	8	39	all
CTSC	GB:AC011088_8	125028	153	A > G	MIS	I>T	153	10	0.9	0	2	8	AFA
CTSC	GB:AC011088_8	125028	153	A > G	MIS	I>T	153	13	1	0	0	13	ASI
CTSC	GB:AC011088_8	125028	153	A > G	MIS	I>T	153	17	0.8235	0	6	11	CAU
CTSC	GB:AC011088_8	125028	153	A > G	MIS	I>T	153	7	1	0	0	7	HIS
CTSC	GB:AC011088_8	150261	149	G > A	MIS	L>F	9	47	0.0106	46	1	0	all
CTSC	GB:AC011088_8	150261	149	G > A	MIS	L>F	9	10	0	10	0	0	AFA
CTSC	GB:AC011088_8	150261	149	G > A	MIS	L>F	9	13	0	13	0	0	ASI
CTSC	GB:AC011088_8	150261	149	G > A	MIS	L>F	9	17	0.0294	16	1	0	CAU
CTSC	GB:AC011088_8	150261	149	G > A	MIS	L>F	9	7	0	7	0	0	HIS
CTSL	GB:AL160279_2	35919	44	C > G	NCD			47	0.0319	44	3	0	all
CTSL	GB:AL160279_2	35919	44	C > G	NCD			10	0	10	0	0	AFA
CTSL	GB:AL160279_2	35919	44	C > G	NCD			13	0	13	0	0	ASI
CTSL	GB:AL160279_2	35919	44	C > G	NCD			17	0.0294	16	1	0	CAU
CTSL	GB:AL160279_2	35919	44	C > G	NCD			7	0.1429	5	2	0	HIS
DAF	GB:AC031978_3	132041	135	C > T	NCD			47	0.0106	46	1	0	all
DAF	GB:AC031978_3	132041	135	C > T	NCD			10	0.05	9	1	0	AFA
DAF	GB:AC031978_3	132041	135	C > T	NCD			13	0	13	0	0	ASI
DAF	GB:AC031978_3	132041	135	C > T	NCD			17	0	17	0	0	CAU
DAF	GB:AC031978_3	132041	135	C > T	NCD			7	0	7	0	0	HIS
DAF	GB:AC031978_3	133893	81	A > G	NCD			47	0.0106	46	1	0	all
DAF	GB:AC031978_3	133893	81	A > G	NCD			10	0.05	9	1	0	AFA
DAF	GB:AC031978_3	133893	81	A > G	NCD			13	0	13	0	0	ASI
DAF	GB:AC031978_3	133893	81	A > G	NCD			17	0	17	0	0	CAU
DAF	GB:AC031978_3	133893	81	A > G	NCD			7	0	7	0	0	HIS
DAF	GB:AC031978_3	133974	162	C > T	SIL	C	98	47	0.0106	46	1	0	all
DAF	GB:AC031978_3	133974	162	C > T	SIL	C	98	10	0.05	9	1	0	AFA
DAF	GB:AC031978_3	133974	162	C > T	SIL	C	98	13	0	13	0	0	ASI
DAF	GB:AC031978_3	133974	162	C > T	SIL	C	98	17	0	17	0	0	CAU
DAF	GB:AC031978_3	133974	162	C > T	SIL	C	98	7	0	7	0	0	HIS
EGF	GB:AC005509	92638	222	C > T	SIL	I	88	46	0.0217	44	2	0	all
EGF	GB:AC005509	92638	222	C > T	SIL	I	88	9	0.1111	7	2	0	AFA
EGF	GB:AC005509	92638	222	C > T	SIL	I	88	13	0	13	0	0	ASI
EGF	GB:AC005509	92638	222	C > T	SIL	I	88	17	0	17	0	0	CAU
EGF	GB:AC005509	92638	222	C > T	SIL	I	88	7	0	7	0	0	HIS
EGF	GB:AC005509	92670	254	A > G	MIS	Q>R	99	46	0.0109	45	1	0	all
EGF	GB:AC005509	92670	254	A > G	MIS	Q>R	99	9	0	9	0	0	AFA
EGF	GB:AC005509	92670	254	A > G	MIS	Q>R	99	13	0	13	0	0	ASI
EGF	GB:AC005509	92670	254	A > G	MIS	Q>R	99	17	0.0294	16	1	0	CAU
EGF	GB:AC005509	92670	254	A > G	MIS	Q>R	99	7	0	7	0	0	HIS

TABLE 2

Gene	Sequence	Seq Offset	Pol Position	DNA change	Qualifier	Peptide change	Codon	No. ind screened	Freq of variant	No. AA	No. AB	No. BB	Ethnic group
EGF	GB:AC005509	92670	85	A > G	MIS	Q>R	99	47	0.0106	46	1	0	all
EGF	GB:AC005509	92670	85	A > G	MIS	Q>R	99	10	0	10	0	0	AFA
EGF	GB:AC005509	92670	85	A > G	MIS	Q>R	99	13	0	13	0	0	ASI
EGF	GB:AC005509	92670	85	A > G	MIS	Q>R	99	17	0.0294	16	1	0	CAU
EGF	GB:AC005509	92670	85	A > G	MIS	Q>R	99	7	0	7	0	0	HIS
EGF	GB:AC005509	92763	178	C > T	NCD			47	0.0638	41	6	0	all
EGF	GB:AC005509	92763	178	C > T	NCD			10	0.1	8	2	0	AFA
EGF	GB:AC005509	92763	178	C > T	NCD			13	0.1538	9	4	0	ASI
EGF	GB:AC005509	92763	178	C > T	NCD			17	0	17	0	0	CAU
EGF	GB:AC005509	92763	178	C > T	NCD			7	0	7	0	0	HIS
EGF	GB:AC005509	94933	97	C > T	MIS	H>Y	151	47	0.0426	43	4	0	all
EGF	GB:AC005509	94933	97	C > T	MIS	H>Y	151	10	0.2	6	4	0	AFA
EGF	GB:AC005509	94933	97	C > T	MIS	H>Y	151	13	0	13	0	0	ASI
EGF	GB:AC005509	94933	97	C > T	MIS	H>Y	151	17	0	17	0	0	CAU
EGF	GB:AC005509	94933	97	C > T	MIS	H>Y	151	7	0	7	0	0	HIS
EGF	GB:AC005509	95444	28	G > C	MIS	D>H	186	46	0.0326	43	3	0	all
EGF	GB:AC005509	95444	28	G > C	MIS	D>H	186	9	0.1667	6	3	0	AFA
EGF	GB:AC005509	95444	28	G > C	MIS	D>H	186	13	0	13	0	0	ASI
EGF	GB:AC005509	95444	28	G > C	MIS	D>H	186	17	0	17	0	0	CAU
EGF	GB:AC005509	95444	28	G > C	MIS	D>H	186	7	0	7	0	0	HIS
EGF	GB:AC005509	96578	69	A > C	NCD			47	0.5	13	21	13	all
EGF	GB:AC005509	96578	69	A > C	NCD			10	0.65	2	3	5	AFA
EGF	GB:AC005509	96578	69	A > C	NCD			13	0.4615	4	6	3	ASI
EGF	GB:AC005509	96578	69	A > C	NCD			17	0.3529	7	8	2	CAU
EGF	GB:AC005509	96578	69	A > C	NCD			7	0.7143	0	4	3	HIS
EGF	GB:AC005509	96660	151	G > C	MIS	D>H	257	47	0.0106	46	1	0	all
EGF	GB:AC005509	96660	151	G > C	MIS	D>H	257	10	0.05	9	1	0	AFA
EGF	GB:AC005509	96660	151	G > C	MIS	D>H	257	13	0	13	0	0	ASI
EGF	GB:AC005509	96660	151	G > C	MIS	D>H	257	17	0	17	0	0	CAU
EGF	GB:AC005509	96660	151	G > C	MIS	D>H	257	7	0	7	0	0	HIS
EGF	GB:AC005509	96842	201	G > A	NCD			47	0.4894	13	22	12	all
EGF	GB:AC005509	96842	201	G > A	NCD			10	0.65	2	3	5	AFA
EGF	GB:AC005509	96842	201	G > A	NCD			13	0.3846	5	6	2	ASI
EGF	GB:AC005509	96842	201	G > A	NCD			17	0.3824	6	9	2	CAU
EGF	GB:AC005509	96842	201	G > A	NCD			7	0.7143	0	4	3	HIS
EGF	GB:AC005509	96853	212	G > A	NCD			47	0.0213	45	2	0	all
EGF	GB:AC005509	96853	212	G > A	NCD			10	0.1	8	2	0	AFA
EGF	GB:AC005509	96853	212	G > A	NCD			13	0	13	0	0	ASI
EGF	GB:AC005509	96853	212	G > A	NCD			17	0	17	0	0	CAU
EGF	GB:AC005509	96853	212	G > A	NCD			7	0	7	0	0	HIS

TABLE 2

Gene	Sequence	Seq Offset	Pol Position	DNA change	Qualifier	Peptide change	Codon	No. ind screened	Freq of variant	No. AA	No. AB	No. BB	Ethnic group
EGF	GB:AC004050	104059	127	GC > GCC	FSH		1135	46	0.0109	45	1	0	all
EGF	GB:AC004050	104059	127	GC > GCC	FSH		1135	9	0	9	0	0	AFA
EGF	GB:AC004050	104059	127	GC > GCC	FSH		1135	13	0.0385	12	1	0	ASI
EGF	GB:AC004050	104059	127	GC > GCC	FSH		1135	17	0	17	0	0	CAU
EGF	GB:AC004050	104059	127	GC > GCC	FSH		1135	7	0	7	0	0	HIS
EGF	GB:AC004050	110749	85	A > G	SIL	L	1063	47	0.0106	46	1	0	all
EGF	GB:AC004050	110749	85	A > G	SIL	L	1063	10	0.05	9	1	0	AFA
EGF	GB:AC004050	110749	85	A > G	SIL	L	1063	13	0	13	0	0	ASI
EGF	GB:AC004050	110749	85	A > G	SIL	L	1063	17	0	17	0	0	CAU
EGF	GB:AC004050	110749	85	A > G	SIL	L	1063	7	0	7	0	0	HIS
EGF	GB:AC005509	112451	117	T > C	SIL	H	365	47	0.0213	45	2	0	all
EGF	GB:AC005509	112451	117	T > C	SIL	H	365	10	0	10	0	0	AFA
EGF	GB:AC005509	112451	117	T > C	SIL	H	365	13	0	13	0	0	ASI
EGF	GB:AC005509	112451	117	T > C	SIL	H	365	17	0.0294	16	1	0	CAU
EGF	GB:AC005509	112451	117	T > C	SIL	H	365	7	0.0714	6	1	0	HIS
EGF	GB:AC005509	113396	37	T > C	NCD			47	0.1596	36	7	4	all
EGF	GB:AC005509	113396	37	T > C	NCD			10	0.3	5	4	1	AFA
EGF	GB:AC005509	113396	37	T > C	NCD			13	0.1538	11	0	2	ASI
EGF	GB:AC005509	113396	37	T > C	NCD			17	0	17	0	0	CAU
EGF	GB:AC005509	113396	37	T > C	NCD			7	0.3571	3	3	1	HIS
EGF	GB:AC005509	113521	88	G > A	MIS	R>K	431	47	0.0851	39	8	0	all
EGF	GB:AC005509	113521	88	G > A	MIS	R>K	431	10	0.05	9	1	0	AFA
EGF	GB:AC005509	113521	88	G > A	MIS	R>K	431	13	0.1923	8	5	0	ASI
EGF	GB:AC005509	113521	88	G > A	MIS	R>K	431	17	0.0588	15	2	0	CAU
EGF	GB:AC005509	113521	88	G > A	MIS	R>K	431	7	0	7	0	0	HIS
EGF	GB:AC005509	114696	39	C > T	NCD			47	0.0638	42	4	1	all
EGF	GB:AC005509	114696	39	C > T	NCD			10	0	10	0	0	AFA
EGF	GB:AC005509	114696	39	C > T	NCD			13	0.2308	8	4	1	ASI
EGF	GB:AC005509	114696	39	C > T	NCD			17	0	17	0	0	CAU
EGF	GB:AC005509	114696	39	C > T	NCD			7	0	7	0	0	HIS
EGF	GB:AC004050	122003	111	A > T	MIS	E>V	920	47	0.9149	0	8	39	all
EGF	GB:AC004050	122003	111	A > T	MIS	E>V	920	10	1	0	0	10	AFA
EGF	GB:AC004050	122003	111	A > T	MIS	E>V	920	13	1	0	0	13	ASI
EGF	GB:AC004050	122003	111	A > T	MIS	E>V	920	17	0.7941	0	7	10	CAU
EGF	GB:AC004050	122003	111	A > T	MIS	E>V	920	7	0.9286	0	1	6	HIS
EGF	GB:AC005509	126323	244	A > G	MIS	I>V	597	46	0.0109	45	1	0	all
EGF	GB:AC005509	126323	244	A > G	MIS	I>V	597	10	0.05	9	1	0	AFA
EGF	GB:AC005509	126323	244	A > G	MIS	I>V	597	12	0	12	0	0	ASI
EGF	GB:AC005509	126323	244	A > G	MIS	I>V	597	17	0	17	0	0	CAU
EGF	GB:AC005509	126323	244	A > G	MIS	I>V	597	7	0	7	0	0	HIS

TABLE 2

Gene	Sequence	Seq Offset	Pol Position	DNA change	Qualifier	Peptide change	Codon	No. ind screened	Freq of variant	No. AA	No. AB	No. BB	Ethnic group
EGF	GB:AC004050	126798	151	C > A	NCD			47	0.0106	46	1	0	all
EGF	GB:AC004050	126798	151	C > A	NCD			10	0.05	9	1	0	AFA
EGF	GB:AC004050	126798	151	C > A	NCD			13	0	13	0	0	ASI
EGF	GB:AC004050	126798	151	C > A	NCD			17	0	17	0	0	CAU
EGF	GB:AC004050	126798	151	C > A	NCD			7	0	7	0	0	HIS
EGF	GB:AC005509	127715	108	C > T	SIL	C	659	47	0.0426	43	4	0	all
EGF	GB:AC005509	127715	108	C > T	SIL	C	659	10	0	10	0	0	AFA
EGF	GB:AC005509	127715	108	C > T	SIL	C	659	13	0	13	0	0	ASI
EGF	GB:AC005509	127715	108	C > T	SIL	C	659	17	0.0588	15	2	0	CAU
EGF	GB:AC005509	127715	108	C > T	SIL	C	659	7	0.1429	5	2	0	HIS
EGF	GB:AC005509	131547	97	A > G	SIL	A	691	45	0.1889	28	17	0	all
EGF	GB:AC005509	131547	97	A > G	SIL	A	691	9	0.3333	3	6	0	AFA
EGF	GB:AC005509	131547	97	A > G	SIL	A	691	13	0.1923	8	5	0	ASI
EGF	GB:AC005509	131547	97	A > G	SIL	A	691	16	0.125	12	4	0	CAU
EGF	GB:AC005509	131547	97	A > G	SIL	A	691	7	0.1429	5	2	0	HIS
EGF	GB:AC005509	131598	148	G > A	MIS	M>I	708	45	0.6222	7	20	18	all
EGF	GB:AC005509	131598	148	G > A	MIS	M>I	708	9	0.8333	0	3	6	AFA
EGF	GB:AC005509	131598	148	G > A	MIS	M>I	708	13	0.6538	1	7	5	ASI
EGF	GB:AC005509	131598	148	G > A	MIS	M>I	708	16	0.4375	6	6	4	CAU
EGF	GB:AC005509	131598	148	G > A	MIS	M>I	708	7	0.7143	0	4	3	HIS
EGF	GB:AC005509	131641	191	G > C	MIS	G>R	723	45	0.0111	44	1	0	all
EGF	GB:AC005509	131641	191	G > C	MIS	G>R	723	9	0	9	0	0	AFA
EGF	GB:AC005509	131641	191	G > C	MIS	G>R	723	13	0	13	0	0	ASI
EGF	GB:AC005509	131641	191	G > C	MIS	G>R	723	16	0	16	0	0	CAU
EGF	GB:AC005509	131641	191	G > C	MIS	G>R	723	7	0.0714	6	1	0	HIS
EGF	GB:AC005509	132511	111	A > T	MIS	D>V	784	47	0.0851	39	8	0	all
EGF	GB:AC005509	132511	111	A > T	MIS	D>V	784	10	0	10	0	0	AFA
EGF	GB:AC005509	132511	111	A > T	MIS	D>V	784	13	0.2308	7	6	0	ASI
EGF	GB:AC005509	132511	111	A > T	MIS	D>V	784	17	0.0588	15	2	0	CAU
EGF	GB:AC005509	132511	111	A > T	MIS	D>V	784	7	0	7	0	0	HIS
G0S2	GB:HS28O10	52341	93	T > C	NCD			44	0.5455	13	14	17	all
G0S2	GB:HS28O10	52341	93	T > C	NCD			9	0.5556	2	4	3	AFA
G0S2	GB:HS28O10	52341	93	T > C	NCD			12	0.4167	5	4	3	ASI
G0S2	GB:HS28O10	52341	93	T > C	NCD			16	0.6875	3	4	9	CAU
G0S2	GB:HS28O10	52341	93	T > C	NCD			7	0.4286	3	2	2	HIS
GADD34	GB:AC026803_2	221941	249	G > A	SIL	R	184	47	0.0106	46	1	0	all
GADD34	GB:AC026803_2	221941	249	G > A	SIL	R	184	10	0.05	9	1	0	AFA
GADD34	GB:AC026803_2	221941	249	G > A	SIL	R	184	13	0	13	0	0	ASI
GADD34	GB:AC026803_2	221941	249	G > A	SIL	R	184	17	0	17	0	0	CAU
GADD34	GB:AC026803_2	221941	249	G > A	SIL	R	184	7	0	7	0	0	HIS

TABLE 2

Gene	Sequence	Seq Offset	Pol Position	DNA change	Qualifier	Peptide change	Codon	No. ind screened	Freq of variant	No. AA	No. AB	No. BB	Ethnic group
IBSP	GB:HUMBNSP03	130	107	G > A				47	0.0106	46	1	0	all
IBSP	GB:HUMBNSP03	130	107	G > A				10	0.05	9	1	0	AFA
IBSP	GB:HUMBNSP03	130	107	G > A				13	0	13	0	0	ASI
IBSP	GB:HUMBNSP03	130	107	G > A				17	0	17	0	0	CAU
IBSP	GB:HUMBNSP03	130	107	G > A				7	0	7	0	0	HIS
IHH	GB:AB018075_1	194	161	G > A	MIS	A>T	167	47	0.0106	46	1	0	all
IHH	GB:AB018075_1	194	161	G > A	MIS	A>T	167	10	0	10	0	0	AFA
IHH	GB:AB018075_1	194	161	G > A	MIS	A>T	167	13	0.0385	12	1	0	ASI
IHH	GB:AB018075_1	194	161	G > A	MIS	A>T	167	17	0	17	0	0	CAU
IHH	GB:AB018075_1	194	161	G > A	MIS	A>T	167	7	0	7	0	0	HIS
IRS1	EM:S85963	850	66	C > T	SIL	D	90	46	0.0761	40	5	1	all
IRS1	EM:S85963	850	66	C > T	SIL	D	90	10	0.2	7	2	1	AFA
IRS1	EM:S85963	850	66	C > T	SIL	D	90	13	0.0385	12	1	0	ASI
IRS1	EM:S85963	850	66	C > T	SIL	D	90	16	0.0313	15	1	0	CAU
IRS1	EM:S85963	850	66	C > T	SIL	D	90	7	0.0714	6	1	0	HIS
IRS1	EM:S85963	1285	262	G > A	SIL	G	235	45	0.1	38	5	2	all
IRS1	EM:S85963	1285	262	G > A	SIL	G	235	9	0	9	0	0	AFA
IRS1	EM:S85963	1285	262	G > A	SIL	G	235	13	0.0385	12	1	0	ASI
IRS1	EM:S85963	1285	262	G > A	SIL	G	235	17	0.1765	12	4	1	CAU
IRS1	EM:S85963	1285	262	G > A	SIL	G	235	6	0.1667	5	0	1	HIS
IRS1	EM:S85963	1285	60	G > A	SIL	G	235	47	0.0426	43	4	0	all
IRS1	EM:S85963	1285	60	G > A	SIL	G	235	10	0.05	9	1	0	AFA
IRS1	EM:S85963	1285	60	G > A	SIL	G	235	13	0.0385	12	1	0	ASI
IRS1	EM:S85963	1285	60	G > A	SIL	G	235	17	0.0294	16	1	0	CAU
IRS1	EM:S85963	1285	60	G > A	SIL	G	235	7	0.0714	6	1	0	HIS
IRS1	EM:S85963	1783	160	T > C	SIL	H	401	45	0.0111	44	1	0	all
IRS1	EM:S85963	1783	160	T > C	SIL	H	401	9	0	9	0	0	AFA
IRS1	EM:S85963	1783	160	T > C	SIL	H	401	13	0.0385	12	1	0	ASI
IRS1	EM:S85963	1783	160	T > C	SIL	H	401	16	0	16	0	0	CAU
IRS1	EM:S85963	1783	160	T > C	SIL	H	401	7	0	7	0	0	HIS
IRS1	EM:S85963	2023	234	C > T	SIL	N	481	47	0.0106	46	1	0	all
IRS1	EM:S85963	2023	234	C > T	SIL	N	481	10	0	10	0	0	AFA
IRS1	EM:S85963	2023	234	C > T	SIL	N	481	13	0	13	0	0	ASI
IRS1	EM:S85963	2023	234	C > T	SIL	N	481	17	0.0294	16	1	0	CAU
IRS1	EM:S85963	2023	234	C > T	SIL	N	481	7	0	7	0	0	HIS
IRS1	EM:S85963	2697	37	G > A	MIS	G>D	706	47	0.0106	46	1	0	all
IRS1	EM:S85963	2697	37	G > A	MIS	G>D	706	10	0	10	0	0	AFA
IRS1	EM:S85963	2697	37	G > A	MIS	G>D	706	13	0.0385	12	1	0	ASI
IRS1	EM:S85963	2697	37	G > A	MIS	G>D	706	17	0	17	0	0	CAU
IRS1	EM:S85963	2697	37	G > A	MIS	G>D	706	7	0	7	0	0	HIS

TABLE 2

Gene	Sequence	Seq Offset	Pol Position	DNA change	Qualifier	Peptide change	Codon	No. ind screened	Freq of variant	No. AA	No. AB	No. BB	Ethnic group
IRS1	EM:S85963	2995	47	A > G	SIL	A	805	47	0.3085	23	19	5	all
IRS1	EM:S85963	2995	47	A > G	SIL	A	805	10	0.45	3	5	2	AFA
IRS1	EM:S85963	2995	47	A > G	SIL	A	805	13	0.4231	4	7	2	ASI
IRS1	EM:S85963	2995	47	A > G	SIL	A	805	17	0.1471	13	3	1	CAU
IRS1	EM:S85963	2995	47	A > G	SIL	A	805	7	0.2857	3	4	0	HIS
IRS1	EM:S85963	3035	87	G > C	MIS	G>R	819	47	0.0106	46	1	0	all
IRS1	EM:S85963	3035	87	G > C	MIS	G>R	819	10	0	10	0	0	AFA
IRS1	EM:S85963	3035	87	G > C	MIS	G>R	819	13	0	13	0	0	ASI
IRS1	EM:S85963	3035	87	G > C	MIS	G>R	819	17	0.0294	16	1	0	CAU
IRS1	EM:S85963	3035	87	G > C	MIS	G>R	819	7	0	7	0	0	HIS
IRS1	EM:S85963	3262	161	G > C	SIL	P	894	47	0.0319	44	3	0	all
IRS1	EM:S85963	3262	161	G > C	SIL	P	894	10	0	10	0	0	AFA
IRS1	EM:S85963	3262	161	G > C	SIL	P	894	13	0	13	0	0	ASI
IRS1	EM:S85963	3262	161	G > C	SIL	P	894	17	0.0588	15	2	0	CAU
IRS1	EM:S85963	3262	161	G > C	SIL	P	894	7	0.0714	6	1	0	HIS
IRS1	EM:S85963	3349	248	G > A	SIL	R	923	47	0.0426	43	4	0	all
IRS1	EM:S85963	3349	248	G > A	SIL	R	923	10	0.05	9	1	0	AFA
IRS1	EM:S85963	3349	248	G > A	SIL	R	923	13	0	13	0	0	ASI
IRS1	EM:S85963	3349	248	G > A	SIL	R	923	17	0.0294	16	1	0	CAU
IRS1	EM:S85963	3349	248	G > A	SIL	R	923	7	0.1429	5	2	0	HIS
IRS1	EM:S85963	3494	155	G > A	MIS	G>R	972	47	0.0957	40	5	2	all
IRS1	EM:S85963	3494	155	G > A	MIS	G>R	972	10	0	10	0	0	AFA
IRS1	EM:S85963	3494	155	G > A	MIS	G>R	972	13	0.0385	12	1	0	ASI
IRS1	EM:S85963	3494	155	G > A	MIS	G>R	972	17	0.1765	12	4	1	CAU
IRS1	EM:S85963	3494	155	G > A	MIS	G>R	972	7	0.1429	6	0	1	HIS
JUN	GB:AL136985_1	8560	66	T > C	SIL	E	303	43	0.0116	42	1	0	all
JUN	GB:AL136985_1	8560	66	T > C	SIL	E	303	10	0.05	9	1	0	AFA
JUN	GB:AL136985_1	8560	66	T > C	SIL	E	303	10	0	10	0	0	ASI
JUN	GB:AL136985_1	8560	66	T > C	SIL	E	303	16	0	16	0	0	CAU
JUN	GB:AL136985_1	8560	66	T > C	SIL	E	303	7	0	7	0	0	HIS
JUN	GB:AL136985_1	8719	225	C > T	SIL	Q	250	43	0.0581	40	1	2	all
JUN	GB:AL136985_1	8719	225	C > T	SIL	Q	250	10	0.25	7	1	2	AFA
JUN	GB:AL136985_1	8719	225	C > T	SIL	Q	250	10	0	10	0	0	ASI
JUN	GB:AL136985_1	8719	225	C > T	SIL	Q	250	16	0	16	0	0	CAU
JUN	GB:AL136985_1	8719	225	C > T	SIL	Q	250	7	0	7	0	0	HIS
KJ_OA11	GB:HS425C14	42204	45	G > A	NCD			46	0.0543	42	3	1	all
KJ_OA11	GB:HS425C14	42204	45	G > A	NCD			9	0	9	0	0	AFA
KJ_OA11	GB:HS425C14	42204	45	G > A	NCD			13	0	13	0	0	ASI
KJ_OA11	GB:HS425C14	42204	45	G > A	NCD			17	0.0588	15	2	0	CAU
KJ_OA11	GB:HS425C14	42204	45	G > A	NCD			7	0.2143	5	1	1	HIS

TABLE 2

Gene	Sequence	Seq Offset	Pol Position	DNA change	Qualifier	Peptide change	Codon	No. Ind screened	Freq of variant	No. AA	No. AB	No. BB	Ethnic group
KJ_OA11	GB:HS425C14	42294	135	G > C	SIL	L	386	46	0.0217	44	2	0	all
KJ_OA11	GB:HS425C14	42294	135	G > C	SIL	L	386	9	0.1111	7	2	0	AFA
KJ_OA11	GB:HS425C14	42294	135	G > C	SIL	L	386	13	0	13	0	0	ASI
KJ_OA11	GB:HS425C14	42294	135	G > C	SIL	L	386	17	0	17	0	0	CAU
KJ_OA11	GB:HS425C14	42294	135	G > C	SIL	L	386	7	0	7	0	0	HIS
KJ_OA11	GB:HS425C14	44297	88	T > C	NCD			47	0.1489	35	10	2	all
KJ_OA11	GB:HS425C14	44297	88	T > C	NCD			10	0.05	9	1	0	AFA
KJ_OA11	GB:HS425C14	44297	88	T > C	NCD			13	0.2692	8	3	2	ASI
KJ_OA11	GB:HS425C14	44297	88	T > C	NCD			17	0.1176	13	4	0	CAU
KJ_OA11	GB:HS425C14	44297	88	T > C	NCD			7	0.1429	5	2	0	HIS
KJ_OA11	GB:HS425C14	55697	54	C > T	NCD			45	0.0111	44	1	0	all
KJ_OA11	GB:HS425C14	55697	54	C > T	NCD			10	0.05	9	1	0	AFA
KJ_OA11	GB:HS425C14	55697	54	C > T	NCD			13	0	13	0	0	ASI
KJ_OA11	GB:HS425C14	55697	54	C > T	NCD			15	0	15	0	0	CAU
KJ_OA11	GB:HS425C14	55697	54	C > T	NCD			7	0	7	0	0	HIS
KJ_OA21	GB:HS453C12	10642	196	A > G	SIL	Y	540	47	0.8191	0	17	30	all
KJ_OA21	GB:HS453C12	10642	196	A > G	SIL	Y	540	10	0.95	0	1	9	AFA
KJ_OA21	GB:HS453C12	10642	196	A > G	SIL	Y	540	13	0.6923	0	8	5	ASI
KJ_OA21	GB:HS453C12	10642	196	A > G	SIL	Y	540	17	0.8235	0	6	11	CAU
KJ_OA21	GB:HS453C12	10642	196	A > G	SIL	Y	540	7	0.8571	0	2	5	HIS
KJ_OA21	GB:HS453C12	11999	79	C > T	SIL	T	457	47	0.2766	26	16	5	all
KJ_OA21	GB:HS453C12	11999	79	C > T	SIL	T	457	10	0.05	9	1	0	AFA
KJ_OA21	GB:HS453C12	11999	79	C > T	SIL	T	457	13	0.2692	6	7	0	ASI
KJ_OA21	GB:HS453C12	11999	79	C > T	SIL	T	457	17	0.4412	6	7	4	CAU
KJ_OA21	GB:HS453C12	11999	79	C > T	SIL	T	457	7	0.2143	5	1	1	HIS
KJ_OA21	GB:HS453C12	16925	27	CG > CCG	NCD			47	0.3085	18	29	0	all
KJ_OA21	GB:HS453C12	16925	27	CG > CCG	NCD			10	0.45	1	9	0	AFA
KJ_OA21	GB:HS453C12	16925	27	CG > CCG	NCD			13	0.1923	8	5	0	ASI
KJ_OA21	GB:HS453C12	16925	27	CG > CCG	NCD			17	0.2941	7	10	0	CAU
KJ_OA21	GB:HS453C12	16925	27	CG > CCG	NCD			7	0.3571	2	5	0	HIS
KJ_OA6	GB:AC005598	132843	24	C > T	SIL	S	48	47	0.0106	46	1	0	all
KJ_OA6	GB:AC005598	132843	24	C > T	SIL	S	48	10	0	10	0	0	AFA
KJ_OA6	GB:AC005598	132843	24	C > T	SIL	S	48	13	0.0385	12	1	0	ASI
KJ_OA6	GB:AC005598	132843	24	C > T	SIL	S	48	17	0	17	0	0	CAU
KJ_OA6	GB:AC005598	132843	24	C > T	SIL	S	48	7	0	7	0	0	HIS
KJ_OA6	GB:AC005598	132878	59	G > A	MIS	R>H	60	47	0.0106	46	1	0	all
KJ_OA6	GB:AC005598	132878	59	G > A	MIS	R>H	60	10	0	10	0	0	AFA
KJ_OA6	GB:AC005598	132878	59	G > A	MIS	R>H	60	13	0	13	0	0	ASI
KJ_OA6	GB:AC005598	132878	59	G > A	MIS	R>H	60	17	0.0294	16	1	0	CAU
KJ_OA6	GB:AC005598	132878	59	G > A	MIS	R>H	60	7	0	7	0	0	HIS

TABLE 2

Gene	Sequence	Seq Offset	Pol Position	DNA change	Qualifier	Peptide change	Codon	No. ind screened	Freq of variant	No. AA	No. AB	No. BB	Ethnic group
KJ_OA6	GB:AC005598	132951	132	C > T	SIL	F	84	47	0.5426	11	21	15	all
KJ_OA6	GB:AC005598	132951	132	C > T	SIL	F	84	10	0.6	3	2	5	AFA
KJ_OA6	GB:AC005598	132951	132	C > T	SIL	F	84	13	0.3462	5	7	1	ASI
KJ_OA6	GB:AC005598	132951	132	C > T	SIL	F	84	17	0.5882	3	8	6	CAU
KJ_OA6	GB:AC005598	132951	132	C > T	SIL	F	84	7	0.7143	0	4	3	HIS
KJ_OA6	GB:AC005598	132967	148	C > T	MIS	P>S	90	47	0.4043	15	26	6	all
KJ_OA6	GB:AC005598	132967	148	C > T	MIS	P>S	90	10	0.3	5	4	1	AFA
KJ_OA6	GB:AC005598	132967	148	C > T	MIS	P>S	90	13	0.5769	1	9	3	ASI
KJ_OA6	GB:AC005598	132967	148	C > T	MIS	P>S	90	17	0.3824	6	9	2	CAU
KJ_OA6	GB:AC005598	132967	148	C > T	MIS	P>S	90	7	0.2857	3	4	0	HIS
KJ_OA6	GB:AC005598	133103	147	G > T	MIS	G>V	135	47	0.0426	45	0	2	all
KJ_OA6	GB:AC005598	133103	147	G > T	MIS	G>V	135	10	0.1	9	0	1	AFA
KJ_OA6	GB:AC005598	133103	147	G > T	MIS	G>V	135	13	0	13	0	0	ASI
KJ_OA6	GB:AC005598	133103	147	G > T	MIS	G>V	135	17	0.0588	16	0	1	CAU
KJ_OA6	GB:AC005598	133103	147	G > T	MIS	G>V	135	7	0	7	0	0	HIS
KJ_OA6	GB:AC005598	133481	188	C > T	NCD			46	0.0109	45	1	0	all
KJ_OA6	GB:AC005598	133481	188	C > T	NCD			9	0	9	0	0	AFA
KJ_OA6	GB:AC005598	133481	188	C > T	NCD			13	0	13	0	0	ASI
KJ_OA6	GB:AC005598	133481	188	C > T	NCD			17	0	17	0	0	CAU
KJ_OA6	GB:AC005598	133481	188	C > T	NCD			7	0.0714	6	1	0	HIS
MMP1	GB:HSU78045	4517	43	A > G	SIL	D	433	47	0.0213	45	2	0	all
MMP1	GB:HSU78045	4517	43	A > G	SIL	D	433	10	0	10	0	0	AFA
MMP1	GB:HSU78045	4517	43	A > G	SIL	D	433	13	0.0769	11	2	0	ASI
MMP1	GB:HSU78045	4517	43	A > G	SIL	D	433	17	0	17	0	0	CAU
MMP1	GB:HSU78045	4517	43	A > G	SIL	D	433	7	0	7	0	0	HIS
MMP1	GB:HSU78045	4661	187	CATG > CG	NCD			47	0.0106	46	1	0	all
MMP1	GB:HSU78045	4661	187	CATG > CG	NCD			10	0.05	9	1	0	AFA
MMP1	GB:HSU78045	4661	187	CATG > CG	NCD			13	0	13	0	0	ASI
MMP1	GB:HSU78045	4661	187	CATG > CG	NCD			17	0	17	0	0	CAU
MMP1	GB:HSU78045	4661	187	CATG > CG	NCD			7	0	7	0	0	HIS
MMP1	GB:HSU78045	4677	203	G > A	NCD			47	0.0638	44	0	3	all
MMP1	GB:HSU78045	4677	203	G > A	NCD			10	0.1	9	0	1	AFA
MMP1	GB:HSU78045	4677	203	G > A	NCD			13	0.1538	11	0	2	ASI
MMP1	GB:HSU78045	4677	203	G > A	NCD			17	0	17	0	0	CAU
MMP1	GB:HSU78045	4677	203	G > A	NCD			7	0	7	0	0	HIS
MMP1	GB:HSU78045	5198	169	A > G	MIS	S>P	382	46	0.0109	45	1	0	all
MMP1	GB:HSU78045	5198	169	A > G	MIS	S>P	382	10	0.05	9	1	0	AFA
MMP1	GB:HSU78045	5198	169	A > G	MIS	S>P	382	13	0	13	0	0	ASI
MMP1	GB:HSU78045	5198	169	A > G	MIS	S>P	382	16	0	16	0	0	CAU
MMP1	GB:HSU78045	5198	169	A > G	MIS	S>P	382	7	0	7	0	0	HIS

TABLE 2

Gene	Sequence	Seq Offset	Pol Position	DNA change	Qualifier	Peptide change	Codon	No. ind screened	Freq of variant	No. AA	No. AB	No. BB	Ethnic group
MMP1	GB:HSU78045	6586	237	T > C	NCD			46	0.413	18	18	10	all
MMP1	GB:HSU78045	6586	237	T > C	NCD			10	0.25	6	3	1	AFA
MMP1	GB:HSU78045	6586	237	T > C	NCD			13	0.7308	0	7	6	ASI
MMP1	GB:HSU78045	6586	237	T > C	NCD			16	0.3125	9	4	3	CAU
MMP1	GB:HSU78045	6586	237	T > C	NCD			7	0.2857	3	4	0	HIS
MMP1	GB:HSU78045	9056	54	C > T	SIL	A	277	46	0.0652	40	6	0	all
MMP1	GB:HSU78045	9056	54	C > T	SIL	A	277	9	0.1111	7	2	0	AFA
MMP1	GB:HSU78045	9056	54	C > T	SIL	A	277	13	0.0385	12	1	0	ASI
MMP1	GB:HSU78045	9056	54	C > T	SIL	A	277	17	0.0882	14	3	0	CAU
MMP1	GB:HSU78045	9056	54	C > T	SIL	A	277	7	0	7	0	0	HIS
MMP1	GB:HSU78045	9120	118	A > G	NCD			46	0.7826	10	0	36	all
MMP1	GB:HSU78045	9120	118	A > G	NCD			9	0.7778	2	0	7	AFA
MMP1	GB:HSU78045	9120	118	A > G	NCD			13	0.8462	2	0	11	ASI
MMP1	GB:HSU78045	9120	118	A > G	NCD			17	0.7647	4	0	13	CAU
MMP1	GB:HSU78045	9120	118	A > G	NCD			7	0.7143	2	0	5	HIS
MMP1	GB:HSU78045	9120	26	A > G	NCD			46	0.0326	43	3	0	all
MMP1	GB:HSU78045	9120	26	A > G	NCD			10	0.1	8	2	0	AFA
MMP1	GB:HSU78045	9120	26	A > G	NCD			12	0	12	0	0	ASI
MMP1	GB:HSU78045	9120	26	A > G	NCD			17	0.0294	16	1	0	CAU
MMP1	GB:HSU78045	9120	26	A > G	NCD			7	0	7	0	0	HIS
MMP1	GB:HSU78045	9126	124	G > A	NCD			46	0.0435	42	4	0	all
MMP1	GB:HSU78045	9126	124	G > A	NCD			9	0	9	0	0	AFA
MMP1	GB:HSU78045	9126	124	G > A	NCD			13	0.0385	12	1	0	ASI
MMP1	GB:HSU78045	9126	124	G > A	NCD			17	0.0294	16	1	0	CAU
MMP1	GB:HSU78045	9126	124	G > A	NCD			7	0.1429	5	2	0	HIS
MMP1	GB:HSU78045	9126	32	G > A	NCD			46	0.0109	45	1	0	all
MMP1	GB:HSU78045	9126	32	G > A	NCD			10	0	10	0	0	AFA
MMP1	GB:HSU78045	9126	32	G > A	NCD			12	0	12	0	0	ASI
MMP1	GB:HSU78045	9126	32	G > A	NCD			17	0	17	0	0	CAU
MMP1	GB:HSU78045	9126	32	G > A	NCD			7	0.0714	6	1	0	HIS
MMP1	GB:HSU78045	9205	111	T > C	NCD			46	0.0109	45	1	0	all
MMP1	GB:HSU78045	9205	111	T > C	NCD			10	0.05	9	1	0	AFA
MMP1	GB:HSU78045	9205	111	T > C	NCD			12	0	12	0	0	ASI
MMP1	GB:HSU78045	9205	111	T > C	NCD			17	0	17	0	0	CAU
MMP1	GB:HSU78045	9205	111	T > C	NCD			7	0	7	0	0	HIS
MMP1	GB:HSU78045	9247	153	T > C	NCD			46	0.0543	42	3	1	all
MMP1	GB:HSU78045	9247	153	T > C	NCD			10	0.15	8	1	1	AFA
MMP1	GB:HSU78045	9247	153	T > C	NCD			12	0.0417	11	1	0	ASI
MMP1	GB:HSU78045	9247	153	T > C	NCD			17	0.0294	16	1	0	CAU
MMP1	GB:HSU78045	9247	153	T > C	NCD			7	0	7	0	0	HIS

TABLE 2

Gene	Sequence	Seq Offset	Pol Position	DNA change	Qualifier	Peptide change	Codon	No. ind screened	Freq of variant	No. AA	No. AB	No. BB	Ethnic group
MMP1	GB:HSU78045	9365	271	G > T	MIS	H>N	228	46	0.0109	45	1	0	all
MMP1	GB:HSU78045	9365	271	G > T	MIS	H>N	228	10	0	10	0	0	AFA
MMP1	GB:HSU78045	9365	271	G > T	MIS	H>N	228	12	0	12	0	0	ASI
MMP1	GB:HSU78045	9365	271	G > T	MIS	H>N	228	17	0.0294	16	1	0	CAU
MMP1	GB:HSU78045	9365	271	G > T	MIS	H>N	228	7	0	7	0	0	HIS
MMP1	GB:HSU78045	11105	177	C > T	SIL	G	105	47	0.0532	43	3	1	all
MMP1	GB:HSU78045	11105	177	C > T	SIL	G	105	10	0.15	8	1	1	AFA
MMP1	GB:HSU78045	11105	177	C > T	SIL	G	105	13	0.0385	12	1	0	ASI
MMP1	GB:HSU78045	11105	177	C > T	SIL	G	105	17	0.0294	16	1	0	CAU
MMP1	GB:HSU78045	11105	177	C > T	SIL	G	105	7	0	7	0	0	HIS
MMP13	GB:AP000789_1	141614	30	C > G	NCD			46	0.0109	45	1	0	all
MMP13	GB:AP000789_1	141614	30	C > G	NCD			9	0.0556	8	1	0	AFA
MMP13	GB:AP000789_1	141614	30	C > G	NCD			13	0	13	0	0	ASI
MMP13	GB:AP000789_1	141614	30	C > G	NCD			17	0	17	0	0	CAU
MMP13	GB:AP000789_1	141614	30	C > G	NCD			7	0	7	0	0	HIS
MMP13	GB:AP000789_1	147095	141	A > G	MIS	H>R	192	43	0.0116	42	1	0	all
MMP13	GB:AP000789_1	147095	141	A > G	MIS	H>R	192	9	0	9	0	0	AFA
MMP13	GB:AP000789_1	147095	141	A > G	MIS	H>R	192	12	0	12	0	0	ASI
MMP13	GB:AP000789_1	147095	141	A > G	MIS	H>R	192	15	0.0333	14	1	0	CAU
MMP13	GB:AP000789_1	147095	141	A > G	MIS	H>R	192	7	0	7	0	0	HIS
MMP13	GB:AP000789_1	157231	114	G > C	MIS	G>R	239	45	0.0111	44	1	0	all
MMP13	GB:AP000789_1	157231	114	G > C	MIS	G>R	239	10	0	10	0	0	AFA
MMP13	GB:AP000789_1	157231	114	G > C	MIS	G>R	239	12	0	12	0	0	ASI
MMP13	GB:AP000789_1	157231	114	G > C	MIS	G>R	239	16	0.0313	15	1	0	CAU
MMP13	GB:AP000789_1	157231	114	G > C	MIS	G>R	239	7	0	7	0	0	HIS
MMP13	GB:AP000789_1	157325	208	A > G	MIS	D>G	270	45	0.0222	44	0	1	all
MMP13	GB:AP000789_1	157325	208	A > G	MIS	D>G	270	10	0.1	9	0	1	AFA
MMP13	GB:AP000789_1	157325	208	A > G	MIS	D>G	270	12	0	12	0	0	ASI
MMP13	GB:AP000789_1	157325	208	A > G	MIS	D>G	270	16	0	16	0	0	CAU
MMP13	GB:AP000789_1	157325	208	A > G	MIS	D>G	270	7	0	7	0	0	HIS
MMP14	GB:AL133448_3	137049	28	TTA > TA	NCD			46	0.9022	0	9	37	all
MMP14	GB:AL133448_3	137049	28	TTA > TA	NCD			10	0.9	0	2	8	AFA
MMP14	GB:AL133448_3	137049	28	TTA > TA	NCD			13	1	0	0	13	ASI
MMP14	GB:AL133448_3	137049	28	TTA > TA	NCD			16	0.8438	0	5	11	CAU
MMP14	GB:AL133448_3	137049	28	TTA > TA	NCD			7	0.8571	0	2	5	HIS
MMP14	GB:AL133448_3	138406	47	G > A	NCD			47	0.0426	43	4	0	all
MMP14	GB:AL133448_3	138406	47	G > A	NCD			10	0.2	6	4	0	AFA
MMP14	GB:AL133448_3	138406	47	G > A	NCD			13	0	13	0	0	ASI
MMP14	GB:AL133448_3	138406	47	G > A	NCD			17	0	17	0	0	CAU
MMP14	GB:AL133448_3	138406	47	G > A	NCD			7	0	7	0	0	HIS

TABLE 2

Gene	Sequence	Seq Offset	Pol Position	DNA change	Qualifier	Peptide change	Codon	No. ind screened	Freq of variant	No. AA	No. AB	No. BB	Ethnic group
MMP14	GB:AL133448_3	138560	201	C > G	SIL	P	259	47	0.0745	40	7	0	all
MMP14	GB:AL133448_3	138560	201	C > G	SIL	P	259	10	0.05	9	1	0	AFA
MMP14	GB:AL133448_3	138560	201	C > G	SIL	P	259	13	0.0769	11	2	0	ASI
MMP14	GB:AL133448_3	138560	201	C > G	SIL	P	259	17	0.0588	15	2	0	CAU
MMP14	GB:AL133448_3	138560	201	C > G	SIL	P	259	7	0.1429	5	2	0	HIS
MMP14	GB:AL133448_3	138653	104	G > A	NCD			47	0.0106	46	1	0	all
MMP14	GB:AL133448_3	138653	104	G > A	NCD			10	0	10	0	0	AFA
MMP14	GB:AL133448_3	138653	104	G > A	NCD			13	0	13	0	0	ASI
MMP14	GB:AL133448_3	138653	104	G > A	NCD			17	0	17	0	0	CAU
MMP14	GB:AL133448_3	138653	104	G > A	NCD			7	0.0714	6	1	0	HIS
MMP14	GB:AL133448_3	139639	101	G > A	MIS	M>I	355	46	0.0217	44	2	0	all
MMP14	GB:AL133448_3	139639	101	G > A	MIS	M>I	355	9	0	9	0	0	AFA
MMP14	GB:AL133448_3	139639	101	G > A	MIS	M>I	355	13	0	13	0	0	ASI
MMP14	GB:AL133448_3	139639	101	G > A	MIS	M>I	355	17	0.0294	16	1	0	CAU
MMP14	GB:AL133448_3	139639	101	G > A	MIS	M>I	355	7	0.0714	6	1	0	HIS
MMP14	GB:AL133448_3	139981	54	C > T	SIL	F	429	47	0.0319	44	3	0	all
MMP14	GB:AL133448_3	139981	54	C > T	SIL	F	429	10	0.1	8	2	0	AFA
MMP14	GB:AL133448_3	139981	54	C > T	SIL	F	429	13	0	13	0	0	ASI
MMP14	GB:AL133448_3	139981	54	C > T	SIL	F	429	17	0	17	0	0	CAU
MMP14	GB:AL133448_3	139981	54	C > T	SIL	F	429	7	0.0714	6	1	0	HIS
MMP14	GB:AL133448_3	139986	59	G > A	MIS	R>H	431	47	0.0106	46	1	0	all
MMP14	GB:AL133448_3	139986	59	G > A	MIS	R>H	431	10	0	10	0	0	AFA
MMP14	GB:AL133448_3	139986	59	G > A	MIS	R>H	431	13	0.0385	12	1	0	ASI
MMP14	GB:AL133448_3	139986	59	G > A	MIS	R>H	431	17	0	17	0	0	CAU
MMP14	GB:AL133448_3	139986	59	G > A	MIS	R>H	431	7	0	7	0	0	HIS
MMP2	GB:AC012182_3	158292	70	G > A	SIL	S	520	46	0.087	40	4	2	all
MMP2	GB:AC012182_3	158292	70	G > A	SIL	S	520	9	0.3889	4	3	2	AFA
MMP2	GB:AC012182_3	158292	70	G > A	SIL	S	520	13	0	13	0	0	ASI
MMP2	GB:AC012182_3	158292	70	G > A	SIL	S	520	17	0	17	0	0	CAU
MMP2	GB:AC012182_3	158292	70	G > A	SIL	S	520	7	0.0714	6	1	0	HIS
MMP2	GB:AC012182_3	163660	71	G > A	NCD			47	0.1702	39	0	8	all
MMP2	GB:AC012182_3	163660	71	G > A	NCD			10	0	10	0	0	AFA
MMP2	GB:AC012182_3	163660	71	G > A	NCD			13	0.3846	8	0	5	ASI
MMP2	GB:AC012182_3	163660	71	G > A	NCD			17	0.0588	16	0	1	CAU
MMP2	GB:AC012182_3	163660	71	G > A	NCD			7	0.2857	5	0	2	HIS
MMP3	EM:HSU78045	52375	108	T > C	SIL	T	400	46	0.0109	45	1	0	all
MMP3	EM:HSU78045	52375	108	T > C	SIL	T	400	10	0.05	9	1	0	AFA
MMP3	EM:HSU78045	52375	108	T > C	SIL	T	400	12	0	12	0	0	ASI
MMP3	EM:HSU78045	52375	108	T > C	SIL	T	400	17	0	17	0	0	CAU
MMP3	EM:HSU78045	52375	108	T > C	SIL	T	400	7	0	7	0	0	HIS

TABLE 2

Gene	Sequence	Seq Offset	Pol Position	DNA change	Qualifier	Peptide change	Codon	No. ind screened	Freq of variant	No. AA	No. AB	No. BB	Ethnic group
MMP3	EM:HSU78045	52411	144	G > A	SIL	I	388	46	0.0109	45	1	0	all
MMP3	EM:HSU78045	52411	144	G > A	SIL	I	388	10	0.05	9	1	0	AFA
MMP3	EM:HSU78045	52411	144	G > A	SIL	I	388	12	0	12	0	0	ASI
MMP3	EM:HSU78045	52411	144	G > A	SIL	I	388	17	0	17	0	0	CAU
MMP3	EM:HSU78045	52411	144	G > A	SIL	I	388	7	0	7	0	0	HIS
MMP3	EM:HSU78045	53771	27	T > A	NCD			47	0.0638	42	4	1	all
MMP3	EM:HSU78045	53771	27	T > A	NCD			10	0	10	0	0	AFA
MMP3	EM:HSU78045	53771	27	T > A	NCD			13	0.0385	12	1	0	ASI
MMP3	EM:HSU78045	53771	27	T > A	NCD			17	0.1471	13	3	1	CAU
MMP3	EM:HSU78045	53771	27	T > A	NCD			7	0	7	0	0	HIS
MMP3	EM:HSU78045	54077	133	C > G	NCD			45	0.0111	44	1	0	all
MMP3	EM:HSU78045	54077	133	C > G	NCD			9	0	9	0	0	AFA
MMP3	EM:HSU78045	54077	133	C > G	NCD			13	0	13	0	0	ASI
MMP3	EM:HSU78045	54077	133	C > G	NCD			16	0.0313	15	1	0	CAU
MMP3	EM:HSU78045	54077	133	C > G	NCD			7	0	7	0	0	HIS
MMP3	EM:HSU78045	54187	81	C > T	NCD			47	0.0106	46	1	0	all
MMP3	EM:HSU78045	54187	81	C > T	NCD			10	0.05	9	1	0	AFA
MMP3	EM:HSU78045	54187	81	C > T	NCD			13	0	13	0	0	ASI
MMP3	EM:HSU78045	54187	81	C > T	NCD			17	0	17	0	0	CAU
MMP3	EM:HSU78045	54187	81	C > T	NCD			7	0	7	0	0	HIS
MMP3	EM:HSU78045	56119	272	C > T	NCD			46	0.1739	37	2	7	all
MMP3	EM:HSU78045	56119	272	C > T	NCD			10	0.2	8	0	2	AFA
MMP3	EM:HSU78045	56119	272	C > T	NCD			13	0.0385	12	1	0	ASI
MMP3	EM:HSU78045	56119	272	C > T	NCD			16	0.2188	12	1	3	CAU
MMP3	EM:HSU78045	56119	272	C > T	NCD			7	0.2857	5	0	2	HIS
MMP3	EM:HSU78045	56507	173	G > C	SIL	T	102	46	0.0217	45	0	1	all
MMP3	EM:HSU78045	56507	173	G > C	SIL	T	102	10	0	10	0	0	AFA
MMP3	EM:HSU78045	56507	173	G > C	SIL	T	102	13	0	13	0	0	ASI
MMP3	EM:HSU78045	56507	173	G > C	SIL	T	102	17	0	17	0	0	CAU
MMP3	EM:HSU78045	56507	173	G > C	SIL	T	102	6	0.1667	5	0	1	HIS
MMP3	EM:HSU78045	56525	191	G > A	SIL	D	96	46	0.5435	1	40	5	all
MMP3	EM:HSU78045	56525	191	G > A	SIL	D	96	10	0.65	0	7	3	AFA
MMP3	EM:HSU78045	56525	191	G > A	SIL	D	96	13	0.5	0	13	0	ASI
MMP3	EM:HSU78045	56525	191	G > A	SIL	D	96	17	0.5588	0	15	2	CAU
MMP3	EM:HSU78045	56525	191	G > A	SIL	D	96	6	0.4167	1	5	0	HIS
MMP3	EM:HSU78045	56680	122	C > T	MIS	E>K	45	47	0.2021	37	1	9	all
MMP3	EM:HSU78045	56680	122	C > T	MIS	E>K	45	10	0.35	6	1	3	AFA
MMP3	EM:HSU78045	56680	122	C > T	MIS	E>K	45	13	0.1538	11	0	2	ASI
MMP3	EM:HSU78045	56680	122	C > T	MIS	E>K	45	17	0.2353	13	0	4	CAU
MMP3	EM:HSU78045	56680	122	C > T	MIS	E>K	45	7	0	7	0	0	HIS

TABLE 2

Gene	Sequence	Seq Offset	Pol Position	DNA change	Qualifier	Peptide change	Codon	No. ind screened	Freq of variant	No. AA	No. AB	No. BB	Ethnic group
MMP9	GB:HUMIVCOL13	87	73	G > A	SIL	V	694	47	0.1915	29	18	0	all
MMP9	GB:HUMIVCOL13	87	73	G > A	SIL	V	694	10	0.15	7	3	0	AFA
MMP9	GB:HUMIVCOL13	87	73	G > A	SIL	V	694	13	0.1923	8	5	0	ASI
MMP9	GB:HUMIVCOL13	87	73	G > A	SIL	V	694	17	0.1765	11	6	0	CAU
MMP9	GB:HUMIVCOL13	87	73	G > A	SIL	V	694	7	0.2857	3	4	0	HIS
MMP9	GB:HUMIVCOL13	132	118	C > T	NCD			47	0.7766	3	15	29	all
MMP9	GB:HUMIVCOL13	132	118	C > T	NCD			10	0.85	0	3	7	AFA
MMP9	GB:HUMIVCOL13	132	118	C > T	NCD			13	0.8077	0	5	8	ASI
MMP9	GB:HUMIVCOL13	132	118	C > T	NCD			17	0.7941	1	5	11	CAU
MMP9	GB:HUMIVCOL13	132	118	C > T	NCD			7	0.5714	2	2	3	HIS
MMP9	GB:HUMIVCOL01	677	111	C > T	MIS	A>V	20	47	0.0106	46	1	0	all
MMP9	GB:HUMIVCOL01	677	111	C > T	MIS	A>V	20	10	0	10	0	0	AFA
MMP9	GB:HUMIVCOL01	677	111	C > T	MIS	A>V	20	13	0.0385	12	1	0	ASI
MMP9	GB:HUMIVCOL01	677	111	C > T	MIS	A>V	20	17	0	17	0	0	CAU
MMP9	GB:HUMIVCOL01	677	111	C > T	MIS	A>V	20	7	0	7	0	0	HIS
MMP9	GB:HUMIVCOL01	677	92	C > T	MIS	A>V	20	47	0.0106	46	1	0	all
MMP9	GB:HUMIVCOL01	677	92	C > T	MIS	A>V	20	10	0	10	0	0	AFA
MMP9	GB:HUMIVCOL01	677	92	C > T	MIS	A>V	20	13	0.0385	12	1	0	ASI
MMP9	GB:HUMIVCOL01	677	92	C > T	MIS	A>V	20	17	0	17	0	0	CAU
MMP9	GB:HUMIVCOL01	677	92	C > T	MIS	A>V	20	7	0	7	0	0	HIS
MSF	GB:AL133553_7	190505	30	A > C	MIS	D>A	220	47	0.0213	45	2	0	all
MSF	GB:AL133553_7	190505	30	A > C	MIS	D>A	220	10	0.05	9	1	0	AFA
MSF	GB:AL133553_7	190505	30	A > C	MIS	D>A	220	13	0	13	0	0	ASI
MSF	GB:AL133553_7	190505	30	A > C	MIS	D>A	220	17	0	17	0	0	CAU
MSF	GB:AL133553_7	190505	30	A > C	MIS	D>A	220	7	0.0714	6	1	0	HIS
MSF	GB:AL133553_7	190559	84	C > T	MIS	T>M	238	47	0.0213	45	2	0	all
MSF	GB:AL133553_7	190559	84	C > T	MIS	T>M	238	10	0	10	0	0	AFA
MSF	GB:AL133553_7	190559	84	C > T	MIS	T>M	238	13	0	13	0	0	ASI
MSF	GB:AL133553_7	190559	84	C > T	MIS	T>M	238	17	0.0588	15	2	0	CAU
MSF	GB:AL133553_7	190559	84	C > T	MIS	T>M	238	7	0	7	0	0	HIS
MSF	GB:AL133553_7	190755	162	G > A	SIL	K	303	43	0.0581	38	5	0	all
MSF	GB:AL133553_7	190755	162	G > A	SIL	K	303	7	0.0714	6	1	0	AFA
MSF	GB:AL133553_7	190755	162	G > A	SIL	K	303	13	0.1154	10	3	0	ASI
MSF	GB:AL133553_7	190755	162	G > A	SIL	K	303	17	0	17	0	0	CAU
MSF	GB:AL133553_7	190755	162	G > A	SIL	K	303	6	0.0833	5	1	0	HIS
MSF	GB:AL133553_7	193235	32	A > G	MIS	N>S	1130	47	0.0106	46	1	0	all
MSF	GB:AL133553_7	193235	32	A > G	MIS	N>S	1130	10	0.05	9	1	0	AFA
MSF	GB:AL133553_7	193235	32	A > G	MIS	N>S	1130	13	0	13	0	0	ASI
MSF	GB:AL133553_7	193235	32	A > G	MIS	N>S	1130	17	0	17	0	0	CAU
MSF	GB:AL133553_7	193235	32	A > G	MIS	N>S	1130	7	0	7	0	0	HIS

TABLE 2

Gene	Sequence	Seq Offset	Pol Position	DNA change	Qualifier	Peptide change	Codon	No. ind screened	Freq of variant	No. AA	No. AB	No. BB	Ethnic group
MSF	GB:AL133553_7	193258	55	A > G	MIS	M>V	1138	47	0.0106	46	1	0	all
MSF	GB:AL133553_7	193258	55	A > G	MIS	M>V	1138	10	0	10	0	0	AFA
MSF	GB:AL133553_7	193258	55	A > G	MIS	M>V	1138	13	0.0385	12	1	0	ASI
MSF	GB:AL133553_7	193258	55	A > G	MIS	M>V	1138	17	0	17	0	0	CAU
MSF	GB:AL133553_7	193258	55	A > G	MIS	M>V	1138	7	0	7	0	0	HIS
NOG	GB:AC005553	145585	63	G > A	SIL	R	206	47	0.0106	46	1	0	all
NOG	GB:AC005553	145585	63	G > A	SIL	R	206	10	0	10	0	0	AFA
NOG	GB:AC005553	145585	63	G > A	SIL	R	206	13	0	13	0	0	ASI
NOG	GB:AC005553	145585	63	G > A	SIL	R	206	17	0.0294	16	1	0	CAU
NOG	GB:AC005553	145585	63	G > A	SIL	R	206	7	0	7	0	0	HIS
NOTCH3	GB:AC004663_1	5567	126	T > C	NCD			47	0.9468	0	5	42	all
NOTCH3	GB:AC004663_1	5567	126	T > C	NCD			10	0.8	0	4	6	AFA
NOTCH3	GB:AC004663_1	5567	126	T > C	NCD			13	1	0	0	13	ASI
NOTCH3	GB:AC004663_1	5567	126	T > C	NCD			17	1	0	0	17	CAU
NOTCH3	GB:AC004663_1	5567	126	T > C	NCD			7	0.9286	0	1	6	HIS
NOTCH3	GB:AC004663_1	21959	43	A > T	NCD			49	0.5204	0	47	2	all
NOTCH3	GB:AC004663_1	21959	43	A > T	NCD			10	0.5	0	10	0	AFA
NOTCH3	GB:AC004663_1	21959	43	A > T	NCD			13	0.5	0	13	0	ASI
NOTCH3	GB:AC004663_1	21959	43	A > T	NCD			17	0.5	0	17	0	CAU
NOTCH3	GB:AC004663_1	21959	43	A > T	NCD			7	0.5	0	7	0	HIS
NOTCH3	GB:AC004663_1	21959	43	A > T	NCD			2					UNK
NOTCH3	GB:AC004663_1	22353	245	C > T	MIS	V>M	1143	47	0.0319	44	3	0	all
NOTCH3	GB:AC004663_1	22353	245	C > T	MIS	V>M	1143	10	0.05	9	1	0	AFA
NOTCH3	GB:AC004663_1	22353	245	C > T	MIS	V>M	1143	13	0	13	0	0	ASI
NOTCH3	GB:AC004663_1	22353	245	C > T	MIS	V>M	1143	17	0	17	0	0	CAU
NOTCH3	GB:AC004663_1	22353	245	C > T	MIS	V>M	1143	7	0.1429	5	2	0	HIS
NOTCH3	GB:AC004663_1	23922	149	C > G	MIS	A>P	980	47	0.0426	43	4	0	all
NOTCH3	GB:AC004663_1	23922	149	C > G	MIS	A>P	980	10	0.1	8	2	0	AFA
NOTCH3	GB:AC004663_1	23922	149	C > G	MIS	A>P	980	13	0	13	0	0	ASI
NOTCH3	GB:AC004663_1	23922	149	C > G	MIS	A>P	980	17	0	17	0	0	CAU
NOTCH3	GB:AC004663_1	23922	149	C > G	MIS	A>P	980	7	0.1429	5	2	0	HIS
NOTCH3	GB:AC004663_1	24045	272	T > C	NCD			47	0.0213	45	2	0	all
NOTCH3	GB:AC004663_1	24045	272	T > C	NCD			10	0.05	9	1	0	AFA
NOTCH3	GB:AC004663_1	24045	272	T > C	NCD			13	0	13	0	0	ASI
NOTCH3	GB:AC004663_1	24045	272	T > C	NCD			17	0	17	0	0	CAU
NOTCH3	GB:AC004663_1	24045	272	T > C	NCD			7	0.0714	6	1	0	HIS
NOTCH3	GB:AC004663_1	28173	175	C > T	MIS	R>H	727	47	0.0106	46	1	0	all
NOTCH3	GB:AC004663_1	28173	175	C > T	MIS	R>H	727	10	0.05	9	1	0	AFA
NOTCH3	GB:AC004663_1	28173	175	C > T	MIS	R>H	727	13	0	13	0	0	ASI
NOTCH3	GB:AC004663_1	28173	175	C > T	MIS	R>H	727	17	0	17	0	0	CAU

TABLE 2

Gene	Sequence	Seq Offset	Pol Position	DNA change	Qualifier	Peptide change	Codon	No. ind screened	Freq of variant	No. AA	No. AB	No. BB	Ethnic group
NOTCH3	GB:AC004663_1	28173	175	C > T	MIS	R>H	727	7	0	7	0	0	HIS
NOTCH3	GB:AC004663_1	28749	174	C > T	MIS	R>H	640	47	0.0426	43	4	0	all
NOTCH3	GB:AC004663_1	28749	174	C > T	MIS	R>H	640	10	0.2	6	4	0	AFA
NOTCH3	GB:AC004663_1	28749	174	C > T	MIS	R>H	640	13	0	13	0	0	ASI
NOTCH3	GB:AC004663_1	28749	174	C > T	MIS	R>H	640	17	0	17	0	0	CAU
NOTCH3	GB:AC004663_1	28749	174	C > T	MIS	R>H	640	7	0	7	0	0	HIS
NOTCH3	GB:AC004663_1	29997	50	G > C	NCD			46	0.022	43	2	0	all
NOTCH3	GB:AC004663_1	29997	50	G > C	NCD			10	0.1	8	2	0	AFA
NOTCH3	GB:AC004663_1	29997	50	G > C	NCD			13	0	13	0	0	ASI
NOTCH3	GB:AC004663_1	29997	50	G > C	NCD			16	0	16	0	0	CAU
NOTCH3	GB:AC004663_1	29997	50	G > C	NCD			7	0	6	0	0	HIS
PDNP1	98092911024828:	201	22	A > G				47	0.0106	46	1	0	all
PDNP1	98092911024828:	201	22	A > G				10	0	10	0	0	AFA
PDNP1	98092911024828:	201	22	A > G				13	0.0385	12	1	0	ASI
PDNP1	98092911024828:	201	22	A > G				17	0	17	0	0	CAU
PDNP1	98092911024828:	201	22	A > G				7	0	7	0	0	HIS
PDNP1	98092910591328	232	78	C > T				47	0.0213	45	2	0	all
PDNP1	98092910591328	232	78	C > T				10	0	10	0	0	AFA
PDNP1	98092910591328	232	78	C > T				13	0	13	0	0	ASI
PDNP1	98092910591328	232	78	C > T				17	0.0588	15	2	0	CAU
PDNP1	98092910591328	232	78	C > T				7	0	7	0	0	HIS
PDNP1	98092911013628:	364	82	C > T				47	0.0106	46	1	0	all
PDNP1	98092911013628:	364	82	C > T				10	0.05	9	1	0	AFA
PDNP1	98092911013628:	364	82	C > T				13	0	13	0	0	ASI
PDNP1	98092911013628:	364	82	C > T				17	0	17	0	0	CAU
PDNP1	98092911013628:	364	82	C > T				7	0	7	0	0	HIS
PLA2G2A	GB:AL358253_1	51295	29	G > A	NCD			47	0.1277	35	12	0	all
PLA2G2A	GB:AL358253_1	51295	29	G > A	NCD			10	0.15	7	3	0	AFA
PLA2G2A	GB:AL358253_1	51295	29	G > A	NCD			13	0.0385	12	1	0	ASI
PLA2G2A	GB:AL358253_1	51295	29	G > A	NCD			17	0.2059	10	7	0	CAU
PLA2G2A	GB:AL358253_1	51295	29	G > A	NCD			7	0.0714	6	1	0	HIS
PLA2G2A	GB:AL358253_1	51364	98	G > A	SIL	Y	44	47	0.1277	35	12	0	all
PLA2G2A	GB:AL358253_1	51364	98	G > A	SIL	Y	44	10	0.15	7	3	0	AFA
PLA2G2A	GB:AL358253_1	51364	98	G > A	SIL	Y	44	13	0.0385	12	1	0	ASI
PLA2G2A	GB:AL358253_1	51364	98	G > A	SIL	Y	44	17	0.2059	10	7	0	CAU
PLA2G2A	GB:AL358253_1	51364	98	G > A	SIL	Y	44	7	0.0714	6	1	0	HIS
PLA2G2A	GB:AL358253_1	51400	134	C > G	SIL	T	32	47	0.0213	45	2	0	all
PLA2G2A	GB:AL358253_1	51400	134	C > G	SIL	T	32	10	0	10	0	0	AFA
PLA2G2A	GB:AL358253_1	51400	134	C > G	SIL	T	32	13	0.0385	12	1	0	ASI
PLA2G2A	GB:AL358253_1	51400	134	C > G	SIL	T	32	17	0.0294	16	1	0	CAU

TABLE 2

Gene	Sequence	Seq Offset	Pol Position	DNA change	Qualifier	Peptide change	Codon	No. ind screened	Freq of variant	No. AA	No. AB	No. BB	Ethnic group
PLA2G2A	GB:AL358253_1	51400	134	C > G	SIL	T	32	7	0	7	0	0	HIS
PLA2G2A	GB:AL358253_1	51454	188	G > A	SIL	G	14	47	0.0106	46	1	0	all
PLA2G2A	GB:AL358253_1	51454	188	G > A	SIL	G	14	10	0	10	0	0	AFA
PLA2G2A	GB:AL358253_1	51454	188	G > A	SIL	G	14	13	0.0385	12	1	0	ASI
PLA2G2A	GB:AL358253_1	51454	188	G > A	SIL	G	14	17	0	17	0	0	CAU
PLA2G2A	GB:AL358253_1	51454	188	G > A	SIL	G	14	7	0	7	0	0	HIS
PLA2G2A	GB:AL358253_1	51503	237	C > T	NCD			47	0.5	1	45	1	all
PLA2G2A	GB:AL358253_1	51503	237	C > T	NCD			10	0.5	0	10	0	AFA
PLA2G2A	GB:AL358253_1	51503	237	C > T	NCD			13	0.5385	0	12	1	ASI
PLA2G2A	GB:AL358253_1	51503	237	C > T	NCD			17	0.4706	1	16	0	CAU
PLA2G2A	GB:AL358253_1	51503	237	C > T	NCD			7	0.5	0	7	0	HIS
PPP1R5	GB:AC020691_2	106523	122	T > G	MIS	D>E	110	47	0.0319	44	3	0	all
PPP1R5	GB:AC020691_2	106523	122	T > G	MIS	D>E	110	10	0.1	8	2	0	AFA
PPP1R5	GB:AC020691_2	106523	122	T > G	MIS	D>E	110	13	0	13	0	0	ASI
PPP1R5	GB:AC020691_2	106523	122	T > G	MIS	D>E	110	17	0	17	0	0	CAU
PPP1R5	GB:AC020691_2	106523	122	T > G	MIS	D>E	110	7	0.0714	6	1	0	HIS
PRSS11	GB:AF157623_1	44762	42	G > T	NCD			47	0.0106	46	1	0	all
PRSS11	GB:AF157623_1	44762	42	G > T	NCD			10	0.05	9	1	0	AFA
PRSS11	GB:AF157623_1	44762	42	G > T	NCD			13	0	13	0	0	ASI
PRSS11	GB:AF157623_1	44762	42	G > T	NCD			17	0	17	0	0	CAU
PRSS11	GB:AF157623_1	44762	42	G > T	NCD			7	0	7	0	0	HIS
PRSS11	GB:AF157623_1	45470	66	C > T	SIL	I	251	47	0.0426	43	4	0	all
PRSS11	GB:AF157623_1	45470	66	C > T	SIL	I	251	10	0.05	9	1	0	AFA
PRSS11	GB:AF157623_1	45470	66	C > T	SIL	I	251	13	0	13	0	0	ASI
PRSS11	GB:AF157623_1	45470	66	C > T	SIL	I	251	17	0.0882	14	3	0	CAU
PRSS11	GB:AF157623_1	45470	66	C > T	SIL	I	251	7	0	7	0	0	HIS
PRSS11	GB:AF157623_1	45587	183	C > T	NCD			47	0.2447	27	17	3	all
PRSS11	GB:AF157623_1	45587	183	C > T	NCD			10	0.15	7	3	0	AFA
PRSS11	GB:AF157623_1	45587	183	C > T	NCD			13	0.3846	5	6	2	ASI
PRSS11	GB:AF157623_1	45587	183	C > T	NCD			17	0.1765	11	6	0	CAU
PRSS11	GB:AF157623_1	45587	183	C > T	NCD			7	0.2857	4	2	1	HIS
PRSS11	GB:AF157623_1	62541	83	G > A	NCD			47	0.0319	44	3	0	all
PRSS11	GB:AF157623_1	62541	83	G > A	NCD			10	0.1	8	2	0	AFA
PRSS11	GB:AF157623_1	62541	83	G > A	NCD			13	0	13	0	0	ASI
PRSS11	GB:AF157623_1	62541	83	G > A	NCD			17	0	17	0	0	CAU
PRSS11	GB:AF157623_1	62541	83	G > A	NCD			7	0.0714	6	1	0	HIS
PRSS11	GB:AF157623_1	62545	87	G > A	NCD			47	0.0106	46	1	0	all
PRSS11	GB:AF157623_1	62545	87	G > A	NCD			10	0	10	0	0	AFA
PRSS11	GB:AF157623_1	62545	87	G > A	NCD			13	0	13	0	0	ASI
PRSS11	GB:AF157623_1	62545	87	G > A	NCD			17	0.0294	16	1	0	CAU

TABLE 2

Gene	Sequence	Seq Offset	Pol Position	DNA change	Qualifier	Peptide change	Codon	No. ind screened	Freq of variant	No. AA	No. AB	No. BB	Ethnic group
PRSS11	GB:AF157623_1	62545	87	G > A	NCD			7	0	7	0	0	HIS
PRSS11	GB:AF157623_1	63355	186	ˉGTTTT > T	NCD			47	0.3085	18	29	0	all
PRSS11	GB:AF157623_1	63355	186	ˉGTTTT > T	NCD			10	0.35	3	7	0	AFA
PRSS11	GB:AF157623_1	63355	186	ˉGTTTT > T	NCD			13	0.2308	7	6	0	ASI
PRSS11	GB:AF157623_1	63355	186	ˉGTTTT > T	NCD			17	0.3235	6	11	0	CAU
PRSS11	GB:AF157623_1	63355	186	ˉGTTTT > T	NCD			7	0.3571	2	5	0	HIS
PTGS2	GB:HUMPTGS2	3050	141	G > C	SIL	V	102	46	0.0543	41	5	0	all
PTGS2	GB:HUMPTGS2	3050	141	G > C	SIL	V	102	10	0	10	0	0	AFA
PTGS2	GB:HUMPTGS2	3050	141	G > C	SIL	V	102	13	0	13	0	0	ASI
PTGS2	GB:HUMPTGS2	3050	141	G > C	SIL	V	102	16	0.1563	11	5	0	CAU
PTGS2	GB:HUMPTGS2	3050	141	G > C	SIL	V	102	7	0	7	0	0	HIS
PTGS2	GB:HUMPTGS2	3090	181	C > T	NCD			46	0.0109	45	1	0	all
PTGS2	GB:HUMPTGS2	3090	181	C > T	NCD			10	0	10	0	0	AFA
PTGS2	GB:HUMPTGS2	3090	181	C > T	NCD			13	0.0385	12	1	0	ASI
PTGS2	GB:HUMPTGS2	3090	181	C > T	NCD			16	0	16	0	0	CAU
PTGS2	GB:HUMPTGS2	3090	181	C > T	NCD			7	0	7	0	0	HIS
PTGS2	GB:HUMPTGS2	3174	265	G > C	NCD			46	0.0217	44	2	0	all
PTGS2	GB:HUMPTGS2	3174	265	G > C	NCD			10	0.05	9	1	0	AFA
PTGS2	GB:HUMPTGS2	3174	265	G > C	NCD			13	0.0385	12	1	0	ASI
PTGS2	GB:HUMPTGS2	3174	265	G > C	NCD			16	0	16	0	0	CAU
PTGS2	GB:HUMPTGS2	3174	265	G > C	NCD			7	0	7	0	0	HIS
PTGS2	GB:HUMPTGS2	3793	177	C > T	SIL	S	132	47	0.0106	46	1	0	all
PTGS2	GB:HUMPTGS2	3793	177	C > T	SIL	S	132	10	0	10	0	0	AFA
PTGS2	GB:HUMPTGS2	3793	177	C > T	SIL	S	132	13	0.0385	12	1	0	ASI
PTGS2	GB:HUMPTGS2	3793	177	C > T	SIL	S	132	17	0	17	0	0	CAU
PTGS2	GB:HUMPTGS2	3793	177	C > T	SIL	S	132	7	0	7	0	0	HIS
PTGS2	GB:HUMPTGS2	3829	213	T > C	SIL	D	144	47	0.0106	46	1	0	all
PTGS2	GB:HUMPTGS2	3829	213	T > C	SIL	D	144	10	0	10	0	0	AFA
PTGS2	GB:HUMPTGS2	3829	213	T > C	SIL	D	144	13	0.0385	12	1	0	ASI
PTGS2	GB:HUMPTGS2	3829	213	T > C	SIL	D	144	17	0	17	0	0	CAU
PTGS2	GB:HUMPTGS2	3829	213	T > C	SIL	D	144	7	0	7	0	0	HIS
PTGS2	GB:HUMPTGS2	5605	110	G > A	NCD			45	0.0111	44	1	0	all
PTGS2	GB:HUMPTGS2	5605	110	G > A	NCD			10	0	10	0	0	AFA
PTGS2	GB:HUMPTGS2	5605	110	G > A	NCD			13	0	13	0	0	ASI
PTGS2	GB:HUMPTGS2	5605	110	G > A	NCD			15	0.0333	14	1	0	CAU
PTGS2	GB:HUMPTGS2	5605	110	G > A	NCD			7	0	7	0	0	HIS
PTGS2	GB:HUMPTGS2	5676	181	ΓATTTT > T	NCD			45	0.0111	44	1	0	all
PTGS2	GB:HUMPTGS2	5676	181	ΓATTTT > T	NCD			10	0	10	0	0	AFA
PTGS2	GB:HUMPTGS2	5676	181	ΓATTTT > T	NCD			13	0	13	0	0	ASI
PTGS2	GB:HUMPTGS2	5676	181	ΓATTTT > T	NCD			15	0.0333	14	1	0	CAU

TABLE 2

Gene	Sequence	Seq Offset	Pol Position	DNA change	Qualifier	Peptide change	Codon	No. ind screened	Freq of variant	No. AA	No. AB	No. BB	Ethnic group
PTGS2	GB:HUMPTGS2	5676	181	TATTTT > T	NCD			7	0	7	0	0	HIS
PTGS2	GB:HUMPTGS2	6249	106	G > A	SIL	V	335	47	0.0106	46	1	0	all
PTGS2	GB:HUMPTGS2	6249	106	G > A	SIL	V	335	10	0	10	0	0	AFA
PTGS2	GB:HUMPTGS2	6249	106	G > A	SIL	V	335	13	0.0385	12	1	0	ASI
PTGS2	GB:HUMPTGS2	6249	106	G > A	SIL	V	335	17	0	17	0	0	CAU
PTGS2	GB:HUMPTGS2	6249	106	G > A	SIL	V	335	7	0	7	0	0	HIS
PTGS2	GB:HUMPTGS2	6444	94	G > A	SIL	L	400	47	0.0106	46	1	0	all
PTGS2	GB:HUMPTGS2	6444	94	G > A	SIL	L	400	10	0	10	0	0	AFA
PTGS2	GB:HUMPTGS2	6444	94	G > A	SIL	L	400	13	0.0385	12	1	0	ASI
PTGS2	GB:HUMPTGS2	6444	94	G > A	SIL	L	400	17	0	17	0	0	CAU
PTGS2	GB:HUMPTGS2	6444	94	G > A	SIL	L	400	7	0	7	0	0	HIS
PTGS2	GB:HUMPTGS2	6453	103	T > C	SIL	H	403	47	0.0106	46	1	0	all
PTGS2	GB:HUMPTGS2	6453	103	T > C	SIL	H	403	10	0.05	9	1	0	AFA
PTGS2	GB:HUMPTGS2	6453	103	T > C	SIL	H	403	13	0	13	0	0	ASI
PTGS2	GB:HUMPTGS2	6453	103	T > C	SIL	H	403	17	0	17	0	0	CAU
PTGS2	GB:HUMPTGS2	6453	103	T > C	SIL	H	403	7	0	7	0	0	HIS
PTGS2	GB:HUMPTGS2	7581	32	T > C	NCD			47	0.0106	46	1	0	all
PTGS2	GB:HUMPTGS2	7581	32	T > C	NCD			10	0.05	9	1	0	AFA
PTGS2	GB:HUMPTGS2	7581	32	T > C	NCD			13	0	13	0	0	ASI
PTGS2	GB:HUMPTGS2	7581	32	T > C	NCD			17	0	17	0	0	CAU
PTGS2	GB:HUMPTGS2	7581	32	T > C	NCD			7	0	7	0	0	HIS
PTGS2	GB:HUMPTGS2	7763	214	T > C	MIS	V>A	511	47	0.0106	46	1	0	all
PTGS2	GB:HUMPTGS2	7763	214	T > C	MIS	V>A	511	10	0.05	9	1	0	AFA
PTGS2	GB:HUMPTGS2	7763	214	T > C	MIS	V>A	511	13	0	13	0	0	ASI
PTGS2	GB:HUMPTGS2	7763	214	T > C	MIS	V>A	511	17	0	17	0	0	CAU
PTGS2	GB:HUMPTGS2	7763	214	T > C	MIS	V>A	511	7	0	7	0	0	HIS
PTGS2	GB:HUMPTGS2	7986	89	C > A	SIL	R	585	47	0.0106	46	1	0	all
PTGS2	GB:HUMPTGS2	7986	89	C > A	SIL	R	585	10	0.05	9	1	0	AFA
PTGS2	GB:HUMPTGS2	7986	89	C > A	SIL	R	585	13	0	13	0	0	ASI
PTGS2	GB:HUMPTGS2	7986	89	C > A	SIL	R	585	17	0	17	0	0	CAU
PTGS2	GB:HUMPTGS2	7986	89	C > A	SIL	R	585	7	0	7	0	0	HIS
PTHR1	GB:HSPTHPRH3	104	29	G > A	SIL	A	72	47	0.0106	46	1	0	all
PTHR1	GB:HSPTHPRH3	104	29	G > A	SIL	A	72	10	0	10	0	0	AFA
PTHR1	GB:HSPTHPRH3	104	29	G > A	SIL	A	72	13	0	13	0	0	ASI
PTHR1	GB:HSPTHPRH3	104	29	G > A	SIL	A	72	17	0.0294	16	1	0	CAU
PTHR1	GB:HSPTHPRH3	104	29	G > A	SIL	A	72	7	0	7	0	0	HIS
PTHR1	GB:HSPTHPRH8	311	207	T > C	SIL	N	463	46	0.6413	6	21	19	all
PTHR1	GB:HSPTHPRH8	311	207	T > C	SIL	N	463	9	0.8333	0	3	6	AFA
PTHR1	GB:HSPTHPRH8	311	207	T > C	SIL	N	463	13	0.5769	3	5	5	ASI
PTHR1	GB:HSPTHPRH8	311	207	T > C	SIL	N	463	17	0.6471	1	10	6	CAU

TABLE 2

Gene	Sequence	Seq Offset	Pol Position	DNA change	Qualifier	Peptide change	Codon	No. ind screened	Freq of variant	No. AA	No. AB	No. BB	Ethnic group
PTHR1	GB:HSPTHPRH8	311	207	T > C	SIL	N	463	7	0.5	2	3	2	HIS
SOD2	EM:S77127	1183	47	C > T	MIS	A>V	16	45	0.6111	7	21	17	all
SOD2	EM:S77127	1183	47	C > T	MIS	A>V	16	10	0.55	2	5	3	AFA
SOD2	EM:S77127	1183	47	C > T	MIS	A>V	16	12	0.9167	0	2	10	ASI
SOD2	EM:S77127	1183	47	C > T	MIS	A>V	16	16	0.5	4	8	4	CAU
SOD2	EM:S77127	1183	47	C > T	MIS	A>V	16	7	0.4286	1	6	0	HIS
SOD2	EM:S77127	1456	209	A > C	NCD			45	0.0111	44	1	0	all
SOD2	EM:S77127	1456	209	A > C	NCD			10	0.05	9	1	0	AFA
SOD2	EM:S77127	1456	209	A > C	NCD			12	0	12	0	0	ASI
SOD2	EM:S77127	1456	209	A > C	NCD			17	0	17	0	0	CAU
SOD2	EM:S77127	1456	209	A > C	NCD			6	0	6	0	0	HIS
SOX9	GB:AC007461_8	62482	48	G > A	SIL	H	169	45	0.0667	39	6	0	all
SOX9	GB:AC007461_8	62482	48	G > A	SIL	H	169	10	0	10	0	0	AFA
SOX9	GB:AC007461_8	62482	48	G > A	SIL	H	169	13	0	13	0	0	ASI
SOX9	GB:AC007461_8	62482	48	G > A	SIL	H	169	16	0.125	12	4	0	CAU
SOX9	GB:AC007461_8	62482	48	G > A	SIL	H	169	6	0.1667	4	2	0	HIS
TIMP1	GB:HS230G1	17434	221	G > A	SIL	I	158	47	0.0532	42	5	0	all
TIMP1	GB:HS230G1	17434	221	G > A	SIL	I	158	10	0	10	0	0	AFA
TIMP1	GB:HS230G1	17434	221	G > A	SIL	I	158	13	0.0769	11	2	0	ASI
TIMP1	GB:HS230G1	17434	221	G > A	SIL	I	158	17	0.0294	16	1	0	CAU
TIMP1	GB:HS230G1	17434	221	G > A	SIL	I	158	7	0.1429	5	2	0	HIS
TIMP1	GB:HS230G1	18389	148	A > G	SIL	F	124	47	0.5106	20	6	21	all
TIMP1	GB:HS230G1	18389	148	A > G	SIL	F	124	10	0.75	2	1	7	AFA
TIMP1	GB:HS230G1	18389	148	A > G	SIL	F	124	13	0.5385	5	2	6	ASI
TIMP1	GB:HS230G1	18389	148	A > G	SIL	F	124	17	0.3235	11	1	5	CAU
TIMP1	GB:HS230G1	18389	148	A > G	SIL	F	124	7	0.5714	2	2	3	HIS
TIMP1	GB:HS230G1	18495	254	C > G	NCD			47	0.5106	20	6	21	all
TIMP1	GB:HS230G1	18495	254	C > G	NCD			10	0.75	2	1	7	AFA
TIMP1	GB:HS230G1	18495	254	C > G	NCD			13	0.5385	5	2	6	ASI
TIMP1	GB:HS230G1	18495	254	C > G	NCD			17	0.3235	11	1	5	CAU
TIMP1	GB:HS230G1	18495	254	C > G	NCD			7	0.5714	2	2	3	HIS
TIMP1	GB:HS230G1	18711	173	G > A	SIL	P	87	47	0.0106	46	1	0	all
TIMP1	GB:HS230G1	18711	173	G > A	SIL	P	87	10	0.05	9	1	0	AFA
TIMP1	GB:HS230G1	18711	173	G > A	SIL	P	87	13	0	13	0	0	ASI
TIMP1	GB:HS230G1	18711	173	G > A	SIL	P	87	17	0	17	0	0	CAU
TIMP1	GB:HS230G1	18711	173	G > A	SIL	P	87	7	0	7	0	0	HIS
TIMP2	GB:U44383_1	155	146	G > A	SIL	S	101	47	0.1064	37	10	0	all
TIMP2	GB:U44383_1	155	146	G > A	SIL	S	101	10	0	10	0	0	AFA
TIMP2	GB:U44383_1	155	146	G > A	SIL	S	101	13	0.0769	11	2	0	ASI
TIMP2	GB:U44383_1	155	146	G > A	SIL	S	101	17	0.1471	12	5	0	CAU

TABLE 2

Gene	Sequence	Seq Offset	Pol Position	DNA change	Qualifier	Peptide change	Codon	No. Ind screened	Freq of variant	No. AA	No. AB	No. BB	Ethnic group
TIMP1	GB:HS230G1	18495	254	C > G	NCD			17	0.3235	11	1	5	CAU
TIMP1	GB:HS230G1	18495	254	C > G	NCD			7	0.5714	2	2	3	HIS
TIMP1	GB:HS230G1	18711	173	G > A	SIL	P	87	47	0.0106	46	1	0	all
TIMP1	GB:HS230G1	18711	173	G > A	SIL	P	87	10	0.05	9	1	0	AFA
TIMP1	GB:HS230G1	18711	173	G > A	SIL	P	87	13	0	13	0	0	ASI
TIMP1	GB:HS230G1	18711	173	G > A	SIL	P	87	17	0	17	0	0	CAU
TIMP1	GB:HS230G1	18711	173	G > A	SIL	P	87	7	0	7	0	0	HIS
TIMP2	GB:U44383_1	155	146	G > A	SIL	S	101	47	0.1064	37	10	0	all
TIMP2	GB:U44383_1	155	146	G > A	SIL	S	101	10	0	10	0	0	AFA
TIMP2	GB:U44383_1	155	146	G > A	SIL	S	101	13	0.0769	11	2	0	ASI
TIMP2	GB:U44383_1	155	146	G > A	SIL	S	101	17	0.1471	12	5	0	CAU
TIMP2	GB:U44383_1	155	146	G > A	SIL	S	101	7	0.2143	4	3	0	HIS
TNFAIP6	GB:AC009311_1	140934	117	G > A	SPL			47	0.0213	45	2	0	all
TNFAIP6	GB:AC009311_1	140934	117	G > A	SPL			10	0.1	8	2	0	AFA
TNFAIP6	GB:AC009311_1	140934	117	G > A	SPL			13	0	13	0	0	ASI
TNFAIP6	GB:AC009311_1	140934	117	G > A	SPL			17	0	17	0	0	CAU
TNFAIP6	GB:AC009311_1	140934	117	G > A	SPL			7	0	7	0	0	HIS
TNFAIP6	GB:AC009311_1	140942	125	A > T	NCD			47	0.0213	45	2	0	all
TNFAIP6	GB:AC009311_1	140942	125	A > T	NCD			10	0.1	8	2	0	AFA
TNFAIP6	GB:AC009311_1	140942	125	A > T	NCD			13	0	13	0	0	ASI
TNFAIP6	GB:AC009311_1	140942	125	A > T	NCD			17	0	17	0	0	CAU
TNFAIP6	GB:AC009311_1	140942	125	A > T	NCD			7	0	7	0	0	HIS
TNFAIP6	GB:AC009311_1	144773	83	A > G	MIS	Q>R	144	47	0.1702	33	12	2	all
TNFAIP6	GB:AC009311_1	144773	83	A > G	MIS	Q>R	144	10	0.25	6	3	1	AFA
TNFAIP6	GB:AC009311_1	144773	83	A > G	MIS	Q>R	144	13	0.1538	9	4	0	ASI
TNFAIP6	GB:AC009311_1	144773	83	A > G	MIS	Q>R	144	17	0.1471	13	3	1	CAU
TNFAIP6	GB:AC009311_1	144773	83	A > G	MIS	Q>R	144	7	0.1429	5	2	0	HIS
TNFAIP6	GB:AC009311_1	148229	47	T > C	NCD			45	0.0778	38	7	0	all
TNFAIP6	GB:AC009311_1	148229	47	T > C	NCD			10	0.1	8	2	0	AFA
TNFAIP6	GB:AC009311_1	148229	47	T > C	NCD			12	0.1667	8	4	0	ASI
TNFAIP6	GB:AC009311_1	148229	47	T > C	NCD			16	0	16	0	0	CAU
TNFAIP6	GB:AC009311_1	148229	47	T > C	NCD			7	0.0714	6	1	0	HIS
TNFAIP6	GB:AC009311_1	148245	63	T > C	NCD			45	0.0778	38	7	0	all
TNFAIP6	GB:AC009311_1	148245	63	T > C	NCD			10	0.1	8	2	0	AFA
TNFAIP6	GB:AC009311_1	148245	63	T > C	NCD			12	0.1667	8	4	0	ASI
TNFAIP6	GB:AC009311_1	148245	63	T > C	NCD			16	0	16	0	0	CAU
TNFAIP6	GB:AC009311_1	148245	63	T > C	NCD			7	0.0714	6	1	0	HIS
TNFRSF11B	GB:E15270_1	503	80	C > T	NCD			47	0.0851	39	8	0	all
TNFRSF11B	GB:E15270_1	503	80	C > T	NCD			10	0.1	8	2	0	AFA
TNFRSF11B	GB:E15270_1	503	80	C > T	NCD			13	0	13	0	0	ASI
TNFRSF11B	GB:E15270_1	503	80	C > T	NCD			17	0.1471	12	5	0	CAU
TNFRSF11B	GB:E15270_1	503	80	C > T	NCD			7	0.0714	6	1	0	HIS
TNFRSF11B	GB:E15270_1	4499	46	C > T	NCD			47	0.0745	40	7	0	all

TABLE 2

Gene	Sequence	Seq Offset	Pol Position	DNA change	Qualifier	Peptide change	Codon	No. ind screened	Freq of variant	No. AA	No. AB	No. BB	Ethnic group
TNFRSF11B	GB:E15270_1	4499	46	C > T	NCD			10	0.05	9	1	0	AFA
TNFRSF11B	GB:E15270_1	4499	46	C > T	NCD			13	0.0769	11	2	0	ASI
TNFRSF11B	GB:E15270_1	4499	46	C > T	NCD			17	0.0882	14	3	0	CAU
TNFRSF11B	GB:E15270_1	4499	46	C > T	NCD			7	0.0714	6	1	0	HIS
TNFRSF11B	GB:E15270_1	4661	159	C > T	SIL	S	176	47	0.0106	46	1	0	all
TNFRSF11B	GB:E15270_1	4661	159	C > T	SIL	S	176	10	0.05	9	1	0	AFA
TNFRSF11B	GB:E15270_1	4661	159	C > T	SIL	S	176	13	0	13	0	0	ASI
TNFRSF11B	GB:E15270_1	4661	159	C > T	SIL	S	176	17	0	17	0	0	CAU
TNFRSF11B	GB:E15270_1	4661	159	C > T	SIL	S	176	7	0	7	0	0	HIS
TNFRSF11B	GB:E15270_1	4749	247	TCTG > TG	NCD			47	0.1702	33	12	2	all
TNFRSF11B	GB:E15270_1	4749	247	TCTG > TG	NCD			10	0.5	2	6	2	AFA
TNFRSF11B	GB:E15270_1	4749	247	TCTG > TG	NCD			13	0	13	0	0	ASI
TNFRSF11B	GB:E15270_1	4749	247	TCTG > TG	NCD			17	0.1471	12	5	0	CAU
TNFRSF11B	GB:E15270_1	4749	247	TCTG > TG	NCD			7	0.0714	6	1	0	HIS
TNFRSF11B	GB:E15270_1	6599	44	G > A	NCD			47	0.0106	46	1	0	all
TNFRSF11B	GB:E15270_1	6599	44	G > A	NCD			10	0.05	9	1	0	AFA
TNFRSF11B	GB:E15270_1	6599	44	G > A	NCD			13	0	13	0	0	ASI
TNFRSF11B	GB:E15270_1	6599	44	G > A	NCD			17	0	17	0	0	CAU
TNFRSF11B	GB:E15270_1	6599	44	G > A	NCD			7	0	7	0	0	HIS
TNFRSF11B	GB:E15270_1	6837	103	G > A	SIL	E	228	43	0.0349	40	3	0	all
TNFRSF11B	GB:E15270_1	6837	103	G > A	SIL	E	228	8	0.125	6	2	0	AFA
TNFRSF11B	GB:E15270_1	6837	103	G > A	SIL	E	228	13	0	13	0	0	ASI
TNFRSF11B	GB:E15270_1	6837	103	G > A	SIL	E	228	15	0.0333	14	1	0	CAU
TNFRSF11B	GB:E15270_1	6837	103	G > A	SIL	E	228	7	0	7	0	0	HIS

Column 1: Gene Name
Column 2: Sequence
Column 3: Position offset in link object
Column 4: Position of the polymorphism
Column 5: DNA change
Column 6: Change type: NCD (non-coding), SIL (silent), STP (stop), FSH (frame shift), MIS (missense).
Column 7: The peptide change as a result of the polymorphism. One letter amino acid abbreviations are used. A * indicates a stop codon.
Column 8: This is the codon number which contains the DNA change.
Column 9: The number is the total number of individuals that have been assigned a genotype for the DNA change.
Column 10: Frequency of variant. This figure is calculated as follows: B alleles/(A+B alleles). The figure is specific to those samples tested.
Column K: Number 11: Total number of individuals with homozygote wild type genotype.
Column 12: Number AB, Total number of individuals with heterozygote genotype.
Column 13: Number BB, Total number of individuals with homozygote mutant genotype.
Column 14: Ethnic group. The different ethnic groups used. CAU= Caucasian, AFC= African, ASI= Asian, HIS= Hispanic, UNK= unknown.

What is claimed is:

1. A method of determining susceptibility of an individual to joint space narrowing and/or osteophyte development and/or joint pain comprising identifying whether the individual has at least one polymorphism in a polynucleotide encoding at least one of the proteins listed in Table 1.

2. The method of claim 1, wherein said proteins listed in Table 1 are selected from the group consisting of bone morphogenic protein 2 (BMP2), cartilage intermediate layer protein (CILP), cartilage oligomeric matrix protein (COMP), tissue inhibitor of metalloproteinase 1 (TIMP1), tetranectin (TNA), matrix metalloproteinase 3 (MMP3), and prostaglandin-endoperoxide synthase 2 (PTGS2).

3. The method of claim 1, wherein the joint space narrowing and/or osteophyte development and/or joint pain is associated with a disease.

4. The method of claim 3 wherein the disease is osteoarthritis.

5. The method of claim 1 where at least one of the polymorphisms is selected from the polymorphisms listed in Table 1.

6. The method of claim 1 comprising contacting a sample from the individual with a specific binding agent for the polymorphism and determining whether the agent binds to the polymorphism.

7. The method of claim 1 where the polymorphism in the polynucleotide is determined for more than one allele of the individual.

8. A method for modulating the susceptibility of an individual to joint space narrowing and/or osteophyte development and/or joint pain, comprising identifying the individual by the method of claim 1 and administering to the individual a composition comprising an effective amount of an agent which modulates said susceptibility.

9. The method of claim 8, wherein the joint space narrowing and/or osteophyte development and/or joint pain is associated with a disease.

10. The method of claim 9 wherein the disease is osteoarthritis.

11. A polynucleotide encoding a protein listed in Table 1 having at least one polymorphism in the polynucleotide selected from the group of polymorphisms listed in Table 1 for the polynucleotide.

12. A fragment of a polynucleotide encoding a protein selected from Table 1 having at least one polymorphism in the fragment selected from the group of polymorphisms listed in Table 1.

13. A fragment of claim 12 having a length of 8 to 100 nucleotides.

14. A fragment of claim 12 having a length of 8 to 30 nucleotides.

15. A fragment of claim 12 having a length of 9 to 15 nucleotides.

16. A method of identifying an agent for modulating susceptibility of an individual to joint space narrowing and/or osteophyte development and/or joint pain comprising:

a) contacting a test agent with a polypeptide or a polynucleotide encoding the polypeptide selected from the list of Table 1 having at least one of the polymorphisms selected from the list of Table 1,

b) determining whether the agent is capable of binding to the polypeptide or polynucleotide encoding the polypeptide, and

c) determining whether the activity or expression of the polypeptide or polynucleotide encoding the polypeptide is modulated.

17. A method of formulating a composition comprising

a) identifying an agent for modulating the susceptibility of an individual to joint space narrowing and/or osteophyte development and/or joint pain by the method of claim 16, and

b) formulating the agent with a carrier or diluent.

18. An agent identified by the method of claim 16.

19. A composition for modulating the susceptibility of an individual to joint space narrowing and/or osteophyte development and/or joint pain comprising an agent according to claim 18 and a carrier.

20. A method comprising using an agent of claim 18 in the manufacture of a medicament for modulating susceptibility to joint space narrowing and/or osteophyte development and/or joint pain.

21. A probe, primer or antibody which is capable of selectively detecting a polymorphism listed in Table 1 which is associated with susceptibility to joint space narrowing and/or osteophyte development and/or joint pain.

22. A vector comprising the polynucleotide of claim 11.

23. A host cell line comprising the vector of claim 22.

24. A nonhuman animal which is transgenic for the polynucleotide of claim 11.

25. A cell line comprising the polynucleotide of claim 11.

26. A method of using a cell line of claim 25 to screen for an agent for diagnosis of an individual having susceptibility to joint space narrowing and/or osteophyte development and/or joint pain.

27. A method of using a nonhuman animal of claim 24 to screen for an agent for diagnosis of an individual having susceptibility to joint space narrowing and/or osteophyte development and/or joint pain.

28. A kit for diagnosis of an individual having susceptibility to joint space narrowing and/or osteophyte development and/or joint pain comprising an agent for detection of the polynucleotide of claim 11.

29. The kit of claim 28 further comprising instruction for use of said agent for detection of said polynucleotide.

30. A kit for diagnosis of an individual having susceptibility to joint space narrowing and/or osteophyte development and/or joint pain comprising an agent for detection of the fragment of a polynucleotide of claim 12.

31. The kit of claim 30 further comprising instructions for use of said agent for detection of said fragment.

32. A kit for diagnosis of an individual having susceptibility to joint space narrowing and/or osteophyte development and/or joint pain comprising the probe, primer or antibody of claim 21.

33. The kit of claim 32 further comprising instructions for use of said probe, primer or antibody.

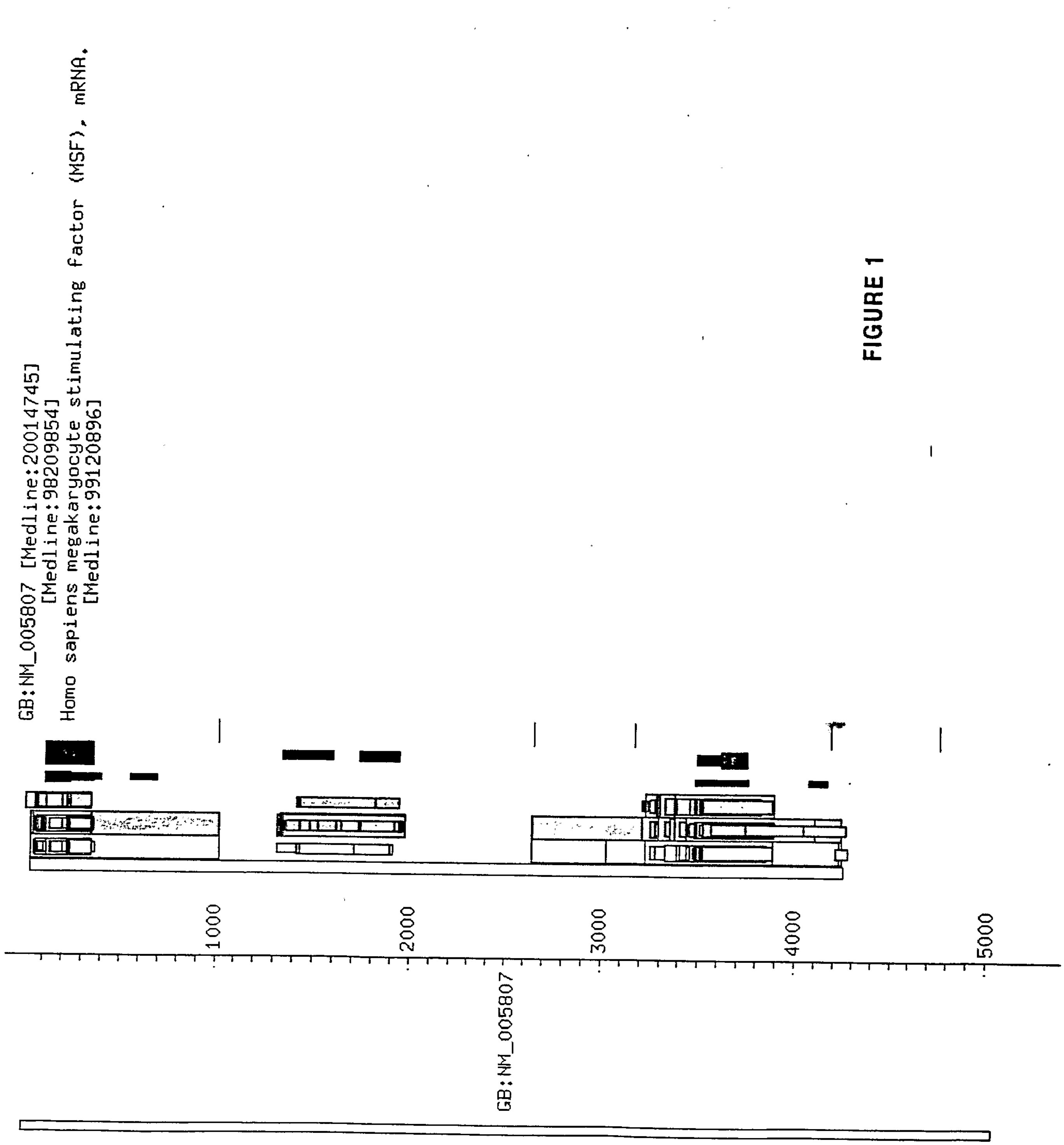

GB:NM_005807 [Medline:20014745]
[Medline:98209854]
Homo sapiens megakaryocyte stimulating factor (MSF), mRNA.
[Medline:99120896]
1000
2000
3000
4000
5000
GB:NM_005807

FIGURE 1

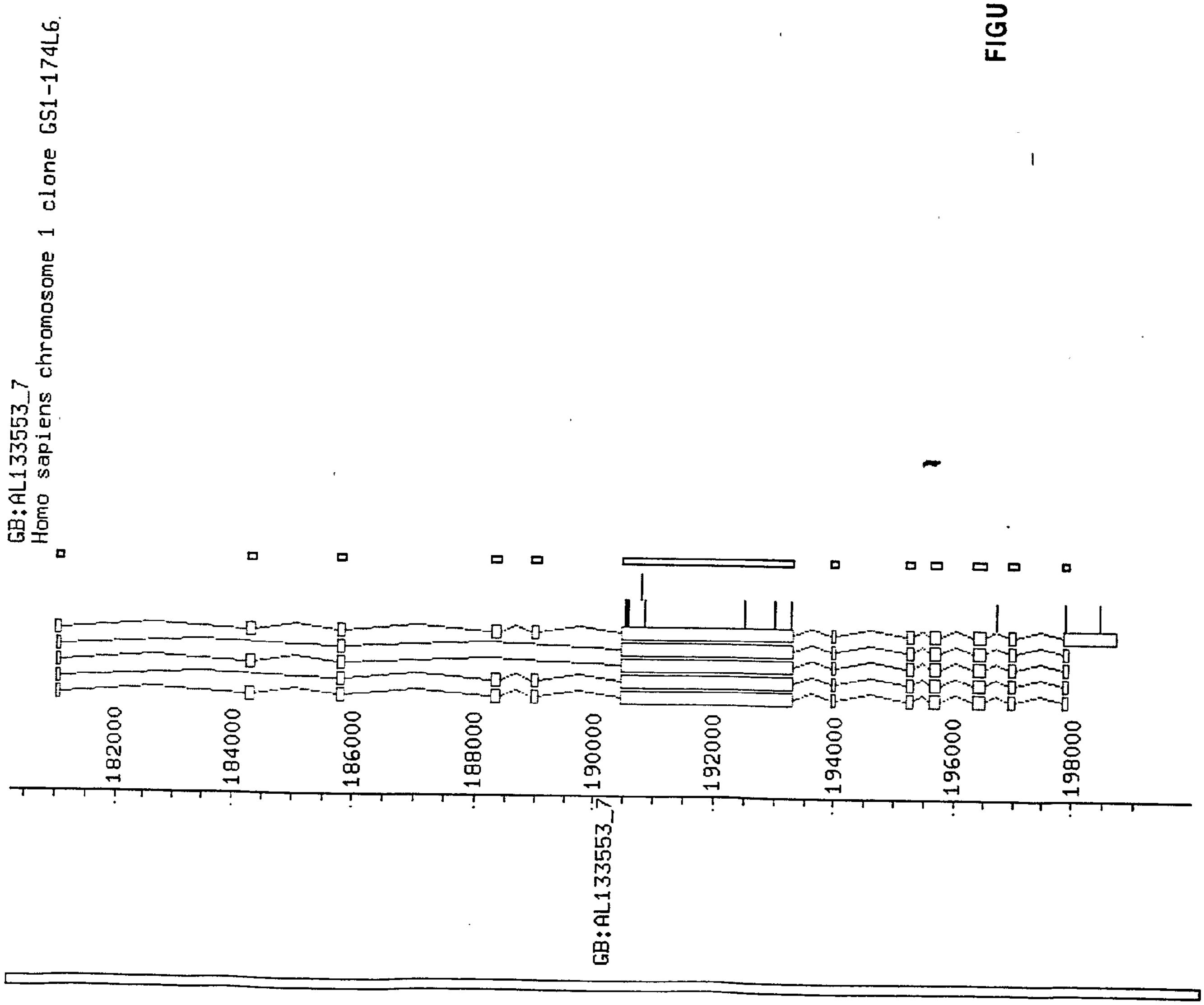

GB:AL133553_7
Homo sapiens chromosome 1 clone GS1-174L6
182000
184000
186000
188000
190000
192000
194000
196000
198000
GB:AL133553_7

FIGURE 2

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 July 2003 (03.07.2003)

PCT

(10) International Publication Number
WO 2003/054166 A3

(51) International Patent Classification[7]: C12Q 1/68, G01N 33/00, 33/53, C12P 19/34, C07H 21/04, C12N 1/21, 5/10, 15/63, A01K 67/00, C07K 14/00

(21) International Application Number: PCT/US2002/041225

(22) International Filing Date: 19 December 2002 (19.12.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/342,603 20 December 2001 (20.12.2001) US

(71) Applicant *(for all designated States except US)*: **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **JONES, Karen, Anne** [GB/GB]; 2 Castle Court, Saffron Walden, Essex, Greater London CB10 1BH (GB). **SCHAFER, Alan** [GB/GB]; 38 Canterbury Street, Cambridge, Cambridgeshire CB43QF (GB).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

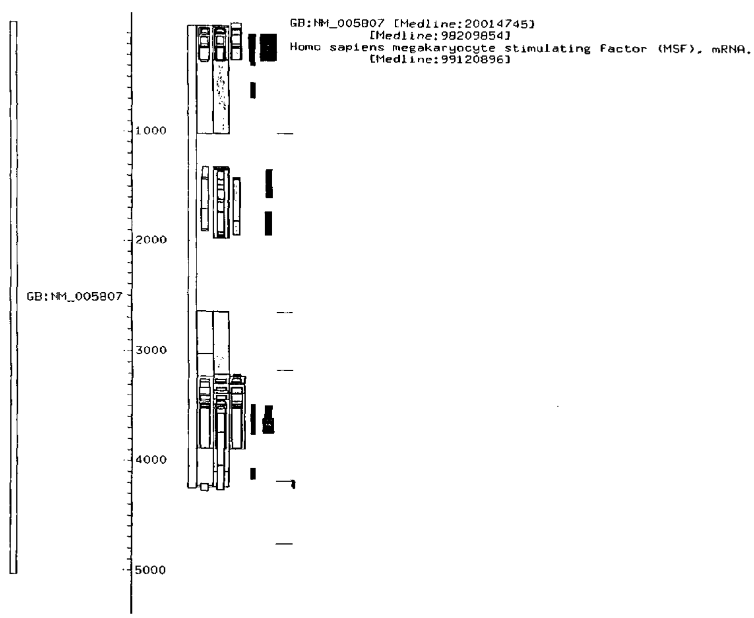
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

[Continued on next page]

(54) **Title:** NUCLEOTIDE POLYMORPHISMS ASSOCIATED WITH OSTEOARTHRITIS



(57) **Abstract:** The invention relates to novel polynucleotides associated with human disease, and in particular to osteoarthritis. The invention further relates to polymorphic polynucleotides associated with osteoarthritis. The invention provides methods of determining if a particular polymorphism predisposes an individual to or is associated with the development of osteoarthritis. The invention also provides methods of detecting the presence of one or more polymorphism as an indicator of osteoarthritis, and provides for use of novel polynucleotides of the invention in the development of drugs and in disease treatment.

WO 2003/054166 A3

(88) Date of publication of the international search report:
18 March 2004

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/41225

A. CLASSIFICATION OF SUBJECT MATTER
IPC(7) : C12Q 1/68; G01N 33/00, 33/53; C12P 19/34; C07H 21/04; C12N 1/21, 5/10, 15/63; A01K 67/00; C07K 14/00
US CL : 435/6, 7.1, 91.2, 252.3, 320.1, 325, 810; 536/23.5, 24.31, 24.33; 530/350; 800/3, 8

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 435/6, 7.1, 91.2, 252.3, 320.1, 325, 810; 536/23.5, 24.31, 24.33; 530/350; 800/3, 8

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	THUR et al. Mutations in cartilage oligomeric matrix protein causing pseudoachondroplasia and multiple epiphyseal dysplasia affect binding of calcium and collagen I, II, and IX. The Journal of Biological Chemistry. 02 March 2001, Vol. 276, No. 9, pages 6083-6092, see entire reference, particularly pages 6083-6084.	11-15, 21-23, 25 ----------- 1-10, 16-20, 24, 26-33
Y	US 6,265,157 B1 (PROCKOP et al) 24 July 2001 (24.07.2001), see entire reference, particularly col 6, line 54-col 10, line 62; col 27, lines 20-38.	1-10, 16-20, 24, 26-33
Y	AHERN, H. Biochemical, reagent kits offer scientists good return on investment. The Scientist. July 1995, Vol. 9, No. 15, pages1-5, see entire reference, particularly page 4/5.	29, 30, 33

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search 28 June 2003 (28.06.2003)	Date of mailing of the international search report 16 JUL 2003
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450 Alexandria, Virginia 22313-1450 Facsimile No. (703)305-3230	Authorized officer Diana B. Johannsen Telephone No. 703/308-0196

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

PCT/US02/41225

Continuation of B. FIELDS SEARCHED Item 3:

USPT, CAPLUS, PGPUB, MEDLINE, LIFESCI, SCISEARCH, EMBASE, BIOSIS, CAPLUS

search terms: bone morphogenic protein 2, bone morphogenetic protein 2, bmp2, cartilage intermediate layer protein, cilp, cartilage oligomeric matrix protein, cartilage matrix protein, cartilage matrix glycoprotein, tissue inhibitor of metalloproteinase 1, timp1, tetranectin, matrix metalloproteinase 3, mmp3, transin, stromelysin, prostaglandin endoperoxide synthase 2, cyclooxygenase 2, cox2, osteoarthritis, joint, multiple epiphyseal dysplasia, polymorphism, mutation, osteophyte; inventors' names